Final Report for AOARD Grant AOARD-114061

Development of Meta Level Communication Analysis using Temporal Data Crystallization and Its Application to Multi Modal Human Communication

22 July 2013

Name of Principal Investigators: Katsumi Nitta

- e-mail address : nitta@dis.titech.ac.jp
- Institution : Tokyo Institute of Technology
- Mailing Address : J2-53, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Japan 226-8502
- Phone : +81-45-924-5214
- Fax : +81-45-924-5214

Period of Performance: 06/22/2011 – 06/21/2013

Abstract:

The purpose of this project is to develop methods to analyze human communication using the *temporal data crystallization* (TDC). We studied following four subjects. At first, we developed a method of meta-level discussion analysis. This method increases precision of TDC analysis by comparing the result of the target record to those of other records. We showed the usefulness of this method by applying it to an example case. Secondly, we studied another method to increase precision of TDC by using the nonverbal information during talking. We showed its effectiveness by applying it to TV debate program. Thirdly, we showed we can extract human relation in detail by analyzing the Tsugo, which is meta level information of human relation such as intension of actions and their constraints. And finally, we showed the way to utilize ubiquitous sensor data such as acceleration sensor and a nearby sensor. We showed a method to extract human relation by applying TDC method to face-to-face contact information.

Introduction:

In the previous two-year AOARD supported project (09-4004), we had developed a text mining method called Temporal Data Crystallization (TDC) which extracts latent interest of opponent from the discussion record. Though TDC detects key utterances which change topics as a form of dummy nodes with a relatively higher degree of accuracy, some dummy nodes are still inadequate ones because the ranking function of TDC is easily affected by noises such as chiming in, repeating what someone said, responding emotionally, and so on. They are sometimes detected as candidates of key utterances by TDC, which reduces the precision of TDC. To apply TDC to the actual discussion records, we need to improve TDC using various kind of information.

The other problem of previous project is that TDC's application is mainly focused on text analysis and more general communication analysis has not been studied so much, though TDC is very general method to analyze communication. Therefore, to show TDC is applicable to analysis of communication analysis, we need to extend TDC and show its effectiveness.

To improve the precision of dummy nodes, and to exploit new application fields, we investigate following subjects.

(1) To develop a meta-level TDC which analyze a discussion record by comparing to other records.
(2) To develop TDC using multi-modal information
(3) To develop a communication analysis method of human relation network by analyzing Tsugo using the same expression as in the abstract.
(4) To develop a communication analysis method using sensory data

Among them, (1) and (2) aim to increase the precision of TDC, and (3) and (4) aim to investigate new application fields of TDC.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
12 JUL 2013	**Final**	**23-06-2011 to 22-06-2013**

4. TITLE AND SUBTITLE
Development of Meta Level Communication Analysis using Temporal Data Crystallization and Its Application to Multi Modal Human Communication

5a. CONTRACT NUMBER
FA23861114061

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Katsumi Nitta

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Computational Science and Systems Science,,Interdisciplinary Graduate School of Engineering and Science,,Tokyo Institute of Technology,4259 Nagatsuta, Midori-ku, Yokohama 176-0024,NA,NA

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AOARD, UNIT 45002, APO, AP, 96338-5002

10. SPONSOR/MONITOR'S ACRONYM(S)
AOARD

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AOARD-114061

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The purpose of this project is to develop methods to analyse human communication using the temporal data crystallization (TDC). The major achievement is the following four. 1) A method of meta-level discussion analysis. This method was shown to increase precision of TDC analysis by comparing the result of the target record to those of other similar records. Its usefulness was shown by applying it to an example case, 2) Another method to increase precision of TDC by using the nonverbal information during talking. Its effectiveness was shown by applying it to TV debate program, 3) Extraction of human relation by analyzing the Tsugo, which is meta level information of human relation such as intension of actions and their constraints, 4) Behavioral analysis of people by utilizing ubiquitous sensor data such as acceleration sensor and a nearby sensor. TDC method was applied to face-to-face contact information and human relation was extracted.

15. SUBJECT TERMS
Brain Science and Engineering; Cognitive Neuroscience; Human-Computer Interface; Intelligent Systems

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**226**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Experiment:

In Section 1, we revisit the idea of Temporal Data Crystallization (TDC) because TDC is a basic technology of this project. In Section 2, we introduce Meta-level Temporal Data Crystallization (Meta-level TDC, in short). Meta-level TDC is a discussion analysis method which observes a discussion record by comparing to another discussion records. In Section 3, we show a TDC method which uses multi-modal information. And also, we show TDC is effective to extract not only key utterances which changed topics but exciting scenes in discussion. In Section 4, we show the Human Tsugo Network (HTN) which analyzes human relation by focusing on constraints between utterances. In Section 5, we show a method to analyze the human communication using several sensors such as an acceleration sensor, nearby sensor, and so on.

1. Temporal Data Crystallization (TDC)

Word Clustering with Temporal Data Crystallization (TDC) is performed as follows. The method proposed by Maeno et al. defines the distance $d(w_i,w_j)$ between each word as the reciprocal of the Jaccard coefficient where the discussion record is considered to be a set of S_1, S_2…, and each utterance S_i is considered to be a set of words that appeared, $\{w_1, w_2 \ldots, w_n\}$,. Next, all words that appeared in utterances are clustered into the given number |C| (C_1, C_2,…,$C_{|C|}$), by utilizing the K-medoids method (Fig. 1.1). When each word is expressed with a node and words having a high Jaccard coefficient are connected with links, a graph that consists of |C| islands (clusters) can be obtained. Each cluster is probably considered to be a single topic.
Next, for each utterance S_i (i=1,2,…), following *ranking functions* $I_{av}(S_i)$ is calculated. Here, $|w_k|$ is the number of appearing of word w_k.

$$I_{av}(S_i) = \frac{1}{|C|}\sum_{j=1}^{|C|} \min_{wk\in(Si\cap Cj)} |w_k| \qquad (1.1)$$

Equation (1.1) is used to find an utterance S_i which contains multiple clusters inside. We select some utterances whose ranking values are relatively high, and for each selected utterance S_k, we insert a dummy node d_k in the graph. The number of selected utterances (the number of dummy nodes) is decided empirically because the proper number depends on features of discussion records. Sometimes we may try several numbers in order to get a suitable set of dummy nodes. The appearance of these dummy nodes suggests that the utterance that corresponds to these nodes refers to several topics. This indicates that other topics are mentioned in the utterance about a certain topic, or a topic is guided to transition to another topic.

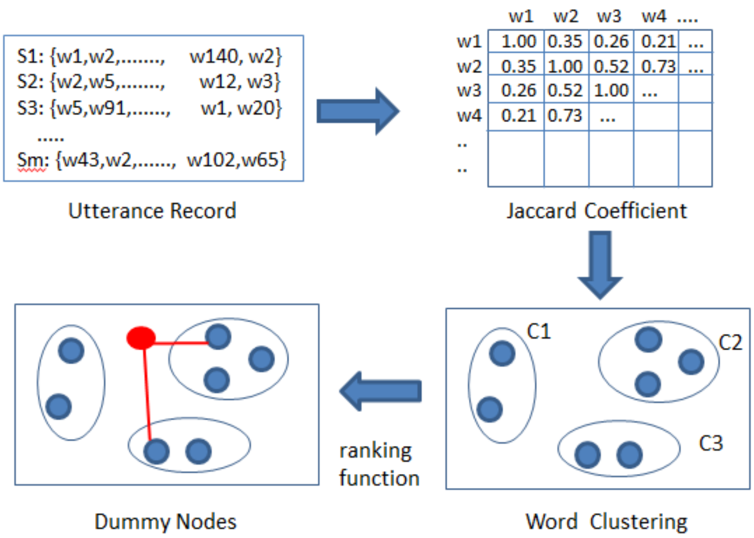


Figure 1.1 Word clustering and dummy nodes

Word clustering shown here is an effective method for analysis of topic transitions based on the discussion record. However, this method has the issue that the clustering precision might decrease when the discussion extends for a long period of time and contains a lot of topics, along with complicated topic transitions.

In such cases, TDC method is used as follows. At first, by applying the word clustering method with DC for the entire discussion record, the words that appear are divided into a given number of clusters (Fig.1.2). Next, a histogram, which shows how words appeared in each cluster as time passed, is obtained. This histogram shown as bar charts indicates each of the clusters. When there is a point where two lines clearly cross, this point is determined to be where topics made a significant shift. Before and after this point, the discussion record is divided into two sections, and then the word clustering method is applied to each of these divided sections. Afterwards, repeating this process divides the discussion record in a hierarchical way.

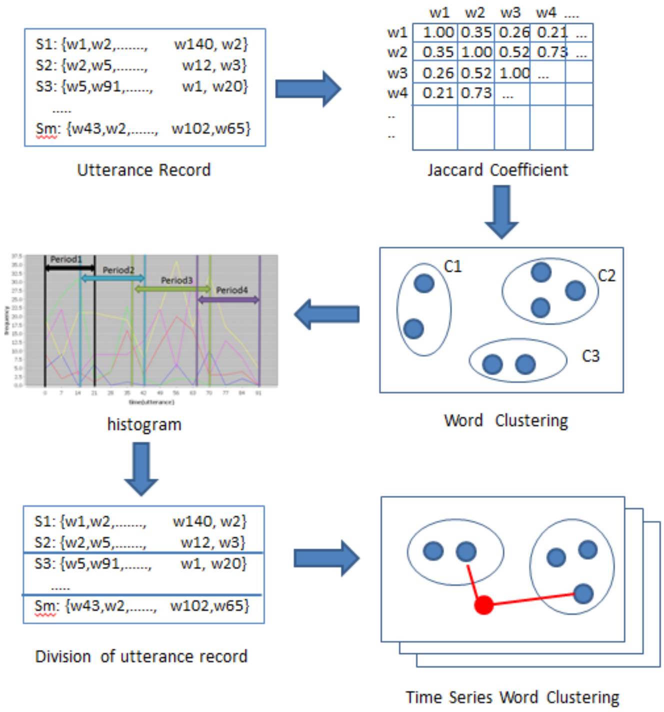


Figure 1.2: Time series word clustering

2. Meta Level Discussion Analysis

In law schools, various discussion trainings such as mock trials, mock mediation and mock negotiation are conducted. In such discussion trainings, students are divided into several groups, and in each group, they discuss the same case problem. Following is an example case.

(Example of case problem)
<u>Muffler Case</u>

Mr. X auctioned off a secondhand automobile muffler on Web-site, and Mr. Y purchased the muffler. Usually, automobile mufflers are made of stainless, but Mr. Y found this muffler was made of a poor material (alster) three month after he purchased it.
On the auction site, Mr. X had not explained about the material, and he had shown only the URL of the Web page of the manufacturer of the muffler. In the Web page, a list of mufflers of the manufacturers is uploaded. But, in the list, all mufflers are made of stainless because the manufacture had stopped making the low quality muffler 4 years ago.

In addition, Mr. X had explained that this muffler is a premium product. Mr. X thought ‘premium’ means a low cost product. However, Mr. Y mistook that premium is a high quality.

So Mr. Y asked Mr. X to cancel the contract and return the money. But, Mr. X rejected it because he hadn't shown false information on the web site, because three month is too long to cancel the contract and because Mr. X had announced that this muffler is "No-claim and No-return" on the auction site.

(End of Example)

After the discussion, their supervisor compares their discussion records and evaluates their discussion skills. In such case, if we analyze each discussion record by TDC, their results (word clusters and dummy nodes) are similar. Therefore, the result of TDC may be improved by comparing to those of other records. For example, even if TDC fails to detect key utterances, they may be corrected by key utterances in the similar scene. We call this method Meta-level TDC.

Meta-level TDC is performed after TDC is applied to each record. At first, the instructor prepares a factor list. A factor is an axiom which describes the content of an issue, a topic, a claim and so on. Followings are examples of factors, which may appear in the discussion of the muffler case. In the muffler case, we prepared 20 factors.

F0: The muffler's material is low quality.

F1: Explanation of the muffler on the Web site is satisfactory.

F2: The catalogue doesn't contain a muffler with low quality material.

F3: By the picture, ordinary people can estimate the material of the muffler.

F4: There is no information about the material of the muffler.

F5: Three months is too late to cancel the contract.

F6: "No Claim, No Return" is the condition of contract.

F7: The seller admit of cancelling the contract.

Then, we attach two types of annotations for each utterance. The first one is the factor appeared in the utterance, and the second one is a label of speech act. A speech act denotes the role of the utterance such as claim (CLAIM), argument (ARG), agreement (AGREE), denial (DENIAL), complement (COMPLEMENT), close-ended-question (CEQ), open-ended-question (OEQ), answer, demand (DEMAND), propose (PROPOSE) and other. Close-ended-question is a form of question, which demands answer from multiple options. And open-ended-question demands answer with free format appropriate to the question such as opinion, argument, agreement, denial, complement, close-ended-question, open-ended-question, answer, demand, propose and other. One utterance may contain one or more speech acts of them. With speech act tags, we are able to handle features in discourse. As a result, each utterance S_i is represented as a speech act label, one or more factors, and a set of words as follows.

S_i = a speech act label + one or more factors + { $w_1, w_2, \ldots, w_n$ }

At the third stage, each discussion record D_j is represented by a vector whose elements are number of factors appeared in the record.

$D_j = (|f_1|, |f_2|, \ldots, |f_n|)^t$ ($|f_k|$ is the number of utterances in which f_k appear)

The similarity between two discussion records is measured by the cosine transformation between two vectors. Table 2.1 is an example of similarity among 8 discussion records of the muffler case. In this table, similarity between record 1 and record 5, and similarity between record 3 and record 7 are relatively high.

Table 2.1 Similarity among discussion records.

	1	2	3	4	5	6	7	8
1	1.0	.65	.67	.59	.87	.69	.77	.77
2		1.0	.59	.56	.78	.76	.68	.71
3			1.0	.75	.81	.61	.85	.80
4				1.0	.74	.57	.83	.62
5					1.0	.64	.83	.81
6						1.0	.71	.73
7							1.0	.78
8								1.0

Table 2.2 is a part of utterances in the record 1 and the record 5. In the Table 2.1, the similarity between these two records is high. However, the sequence of topics is very different as in Table 2.2. The utterances 9 and 10 in the record 1 correspond to utterances 25 and 26 in the record 5 though their presentations are different.

(Record1)

Utterance 9 “The muffler is made of Alster. Usually, the standard material is stainless. I want to return the muffler.”

Utterance 10 “I don’t accept cancel of the contract because three month passed after I sent you the muffler.”

(Record 5)

Utterance 25 “As the muffler is low quality, I wish to cancel the contract.”

Utterance 26 “The contracted is completed because you didn’t claim just after I sent you the muffler.”

In the record 1, TDC selected the utterance 9 as a dummy node. On the contrary, in the record 5, TDC didn’t select the utterance 25 as dummy node. In such case, utterance 25 may be selected as a dummy node because the record 1 and record 5 are similar.

Table 2.2 Utterance sequence of record1 and record 5

Utterance ID	Record1		Record5	
	speaker	speech act + factor	speaker	speech act + factor
1	Y	CLAIM F0 PROPOSE F7	Y	CLAIM F0
2	M		M	CEQ
3	X	REFUSE F7 ARG F1	Y	ANSWER F0
4	M	CEQ	M	OEQ
5	X	ANSWER	X	ANSWER F5
6	M	CEQ	M	
7	X	ANSWER	X	CLAIM F0
8	M		M	CEQ
9	Y	ARG F7<-F0	X	ANSWER
10	X	DENY F7 ARG ¬F7<-F5	M	
◦ ◦ ◦				
25	X	CLAIM F1	Y	ARG F7<-F0
26	Y	DENY F1	X	DENY F7 ARG ¬F7<-F5
◦ ◦ ◦				

3. TDC by Considering Multi-Modal Data (Discussion analysis using Multi-modal information)

In this section, we propose the method of TDC with nonverbal information. To evaluate the effect of this method in the mediation, we need the movie data of the mediation. However, it is hard to obtain

the high quality movie data, because the players of the mock mediation don't show the emotional reactions so much, and because it is prohibited to record the real mediation. Therefore, as a preliminary study, we applied this new method to TV discussion programs where discussants talk passionately. In this program, 14 discussants are selected from journalists, statesmen or commentators, and they discussed several political managements of Japanese Government for 4 hours. The subjects discussed were the relocation of the U.S. air base in Okinawa, economic stimulus measures, and the consumption tax.

3.1 Recognition of Topic Transition using Gesture Information

Our target record is obtained from a discussion where each participant sits in a chair. From discussion video records, we observed salient characteristic of speakers (Table 3.1) and labeled each utterance (Table 3.2).

We show a method for extracting topic transitions. The label of gesture information $\{a_1, a_2, \dots, a_n\}$ is an attribute of the dummy word d_i (equation 3.1).

$$S_i = \{w_1, w_2, \dots, w_m, d_i\} \tag{3.1}$$

Table 3.1 Gesture Labels

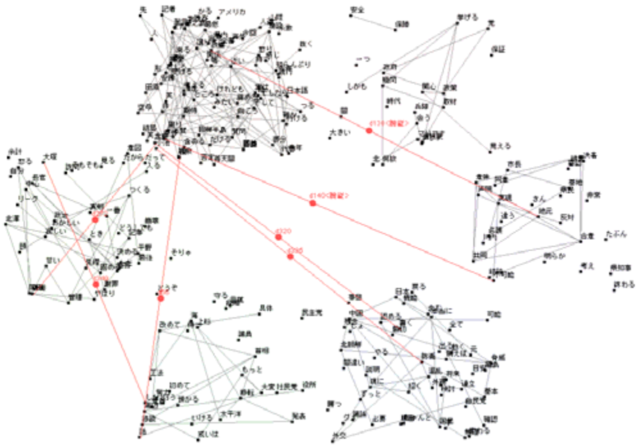
Body part	Label	Meaning of the label
Head	Downward	Looking down
	Forward	Putting the head forward
	Nodding	Nodding the head
Trunk	Rightward	Tilting the trunk to the right
	Backward	Tilting the trunk backward
	Leftward	Tilting the trunk to the left
	trunk Forward	Tilting the trunk forward
	Back and forth	Tilting the trunk back and forth
Hands and arms	Hands horizontal	Moving the hands horizontally
	Hands vertical	Moving the hands vertically
	Folding	Folding the arms
	Bringing together	Bringing hands together
Voice	Loud	Speaking with a loud voice

Table 3.2 Gesture of Discussants

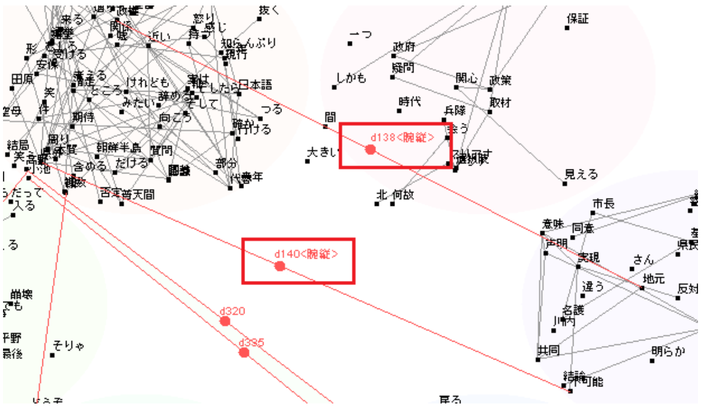
Speaker	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Downward	0	0	0	0	0	0	2	1	0	0	0	0	4	1
Forward	3	1	0	2	9	2	1	1	0	5	0	0	1	0
Nodding	0	0	0	0	1	1	6	2	0	0	0	1	0	1
Rightward	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Backward	0	0	0	0	0	0	0	0	0	2	0	0	1	4
Leftward	0	0	0	0	0	0	0	0	0	1	0	1	0	0
trunk Forward	9	1	4	1	1	1	2	4	0	24	1	0	4	3
Back and forth	1	0	0	0	0	1	0	0	0	0	1	0	0	0
Hands horizontal	0	0	0	2	0	0	3	0	0	0	0	2	1	0
Hands vertical	0	0	0	2	0	3	13	10	3	19	0	3	4	5
Folding	3	2	0	7	0	0	0	0	0	0	0	0	1	3
Bringing together	9	0	0	18	0	0	2	0	0	0	2	2	0	0
Loud	1	0	0	0	0	0	1	0	0	17	0	0	0	0
Total	26	4	4	33	11	8	30	18	3	68	4	9	16	17

After we generate dummy nodes by TDC method, we select the ones with gesture labels. Selected dummy nodes are more reliable than original dummy nodes because key utterances often accompanied by gestures.

Fig. 3.1(a) shows the clustering result of the first half of the show. Black nodes indicate words, while red links indicate dummy nodes. In this Fig., the following topics were shown: the domestic relocation of the air base (Fig. 3.1(a): Upper left), the overseas relocation of the air base (Fig. 3.1(a): Upper right), about the Prime Minister's Office (Fig. 3.1(a): Left), willingness of the local residents (Fig.



(a) Word Clustering by TDC



(b) A magnified graph centered on dummy nodes d138 and d140

Fig. 3.1 Example of TDC with nonverbal information

Table 3.3 Statements (Statements ID136 – ID141)

ID	Speaker	Content
136	Kawauchi	(Omitted) I'm saying that it is impossible to realize this plan. <Crossing his arms>
137	Tawara	This is a bit difficult to understand, we need interpretation. Um, Mr. Otsuka, what is he saying?
138	Otsuka	(Omitted) No consensus has been reached with the local residents, so it means there is no guarantee yet that the scenario goes just according to what was claimed in today's joint declaration. <Arms vertical>
139	Tawara	No, not at all.
140	Mogi	(Omitted) So, the Prime Minister said that, right? Saying what is unrealizable, he also said that at least the air base would be relocated outside Okinawa, during the election campaign. After all, this relocation was impossible. Now he said the base

		would go to Henoko. It's too late to refer to another destination like Henoko, it's totally impossible to relocate the base there. (Omitted) <Arms vertical>
141	Yamagiwa	(Omitted) It does not necessarily mean that all are opposed to the presence of the base. Not all. (Omitted)

3.1(a): Right), about Prime Minister Hatoyama (Fig. 3.1(a): Lower left), and about the deterrent force (Fig. 3.1(a): Lower right). Here, Fig 3.1(b) shows a magnified graph centered on dummy nodes d138 and d140. The attribute "Arms Vertical" (swinging of the arms vertically) was given to statements ID138 and ID140. In Table 3.3, in statement ID138, the topic transitioned from the relocation of the air base to the willingness of the local residents. In statement 140, the topic transitioned from willingness of the local residents to the relocation of the air base. This is an example where topic transitions could be detected by giving gesture labels.

We compared the existing method (TDC only) with the proposed method (TDC with non-verbal information) to examine the accuracy of topic transition. Table 3.4 shows the experimental results using equation 3.2. Here, *num(dummy_node_with_gesture)* means the number of selected dummy nodes with gestures, *num(correct_answer)* means the number of the topic transitions in the selected dummy nodes with gestures, and *num(gesture_label_in_the_topic_transition)* means the number of utterances with gestures and caused topic transition. Here, we must pay attention that equation 3.2 don't consider the number of utterances without gestures. About two thirds of utterances are accompanied with gestures.

$$\Pr ecision = \frac{num(correct\_answers)}{num(dummy\_node\_with\_gestures)}$$
$$\operatorname{Re} call = \frac{num(correct\_answers)}{num(gesture\_label\_in\_the\_topic\_transition)} \tag{3.2}$$

Table 3.4 Detection of Topic

Dummy nodes	TDC			TDC with nonverbal info.		
	Precision	Recall	F-measure	Precision	Recall	F-measure
20	0.15	0.08	0.10	0.30	0.50	0.38
40	0.22	0.24	0.23	0.30	1.00	0.46

According to Table 3.4, the proposed method shows better results in both precision and recall than existing methods.

3.2 Recognition of Exciting Scene (Heat Up Scene) Using Gesture Information

Gesture Information is useful for not only detecting key utterances but detecting exciting scene during the discussion. When utterances with gestures are repeated a few times, we assume such periods correspond to the exciting scene. To confirm this assumption, we conducted an experiment using the same discussion record of the TV debate program.

We extracted places where one of followings are repeated 3 times.

1) the utterances with gestures,
2) the dummy nodes by TDC
3) the dummy nodes with gestures

And then we examined if such scenes correspond to exciting ones or not. Table 3.5 is the result of this experiment.

Table 3.5 Precision of Recognition of Exciting Scene

Utterance with	Dummy	Dummy nodes

gestures	nodes	with gestures
0.52	0.40	0.68

This table shows that precision of the dummy nodes with gestures is higher than other methods.

3.3 Recognition of Topic Transition using Speaker Pairs Information

Next, to improve the accuracy of detecting topic changes, we focused on the change of *speaker pairs*. A *speaker pair* is defined as two persons speaking alternately. For example, let speakers A, B, C and D speak as follows.

A B A B A B A C A D …

We assume that while two persons (A and B) speak alternately, the topic change doesn't occur, and when a change in speaker (C) occurs, the topic may change near this point.

Fig. 3.2 shows that there is a relation between a change in speaker pairs and topic transition. The discussion record is the same as in the previous section. This result shows a 45% of the total topic transition was seen when the speaker pair changed. And, more than 90% topic transitions were within one utterances of the change of speaker pairs in case of this discussion record. Furthermore, a topic transition occurred within six utterances. Therefore we targeted the utterances within the six utterances of the change of speaker pairs for the discovering topic transitions.

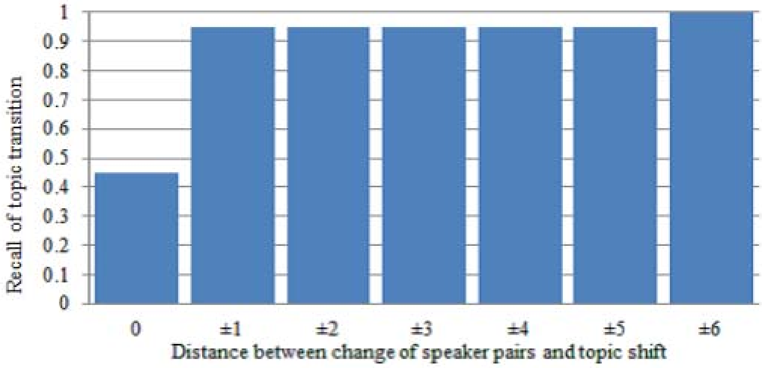


Fig. 3.2: Relation between speaker and Topic Change

We verified that improved extraction accuracy of topic transition using speaker pair information could be achieved (Fig. 3.3). In this figure, we changed the number of selected dummy nodes from 10 to 80, and observed Precision, Recall and F-measure for each. We showed the proposed method (dummy nodes + speaker pair) is more effective than existing method (only dummy node).

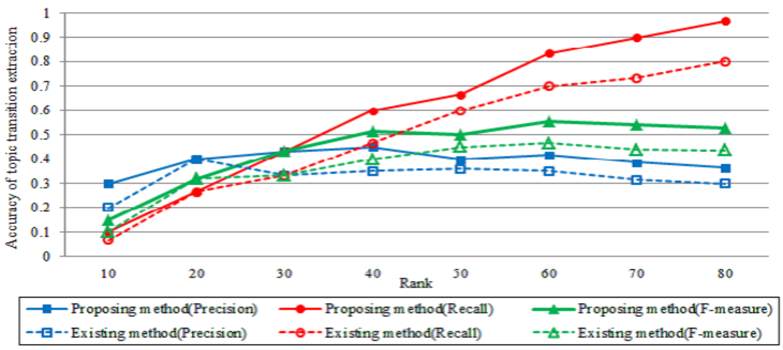


Fig. 3.3 Comparison of Precision and Recall

In this section, we showed that improved extraction accuracy of topic transition using not only text data, but also non-verbal information was achieved (Gesture and Speaker pair).

4. Human Tsugo Network (HTN) reviewed

As given in the literature (Ohsawa et al 2010, Ohsawa and Nishihara 2012), Human Tsugo Network (HTN) is a method to visualize relationships among stakeholders' *tsugoes*. Here, a tsugo means a set {intention (*int*), pre (*pre*)-constraint, post-constraint (*pst*)} that tend to be hidden behind actions and are not likely to be verbalized. Note this notion is quite similar to elements of arguments as proposed in the studies on logical argumentation [Atkinson, Bench-capon, and McBurney 2006]. However, the point of tsugo is that its elements are normally not externalized in humans' conversations whereas elements corresponding to intention, pre-, and post-constraints are explicitly considered in "attacks" to an argument in argumentation studies.

Due to this strong tendency to be hidden, the word *tsugo* has been used in the traditional business communications in Japan, difficult to translate into other languages. For example, "I go home due to a tsugo," without mentioning what really the tsugo is, means one is about to go home due to an intention or a pre-constraint, or considering a post-constraint that is not easy to verbalize – one should do so because of an affair difficult to explain. However, such a secret sometimes causes a delay of businesses due to conflicts that appears later when the hidden elements of tsugoes come out to be revealed. People living in social relationships should externalize one's and others' tsugoes for choosing an action that is admissible, i.e., for realizing an intention under constraints some of which emerge on the way of the action itself.

Toward designing satisfactory products or services, to understand intentions and constraints of stakeholders is an essential step (Goldratt 1987, Carrol 2000, Kushiro and Ohsawa 2006). However, quite essential parts of such information are hidden, and difficult to tell in a simple interview, like the tacit dimension of knowledge (Polanyi 1966, 2009) hidden behind activities of humans in the real life. The tacit dimension should be and can be externalized for and by enabling a creative process of collaboration (Nonaka 1994). Our assumption in tsugology, that is the studies about tsugoes and their applications, is that such a dynamic approach as by Nonaka is desired, because latent intentions and constraints should interact dynamically when they collaborate. That is, we reflect the of dynamics of post-constraints, that may emerge from an action and may make a pre-constraint on others actions, in contrast to static constraints in the literature of design methods.For example, if Mr. X builds a tall building on the intention to expand business, the pre-constraint of habitants in the neighborhood, such as their requirement for sunshine, may be violated. In this case, the disturbance of sunshine is a post-constraint for Mr. X that had not been considered until being externalized due to discussing the plan of building. By noticing such a constraint, Mr.X should think of a new action, such as networking members' distributed hometowns rather than building a new workspace, for realizing his real and latent intention to expand the residential areas of customers. This intention may also get externalized via speaking out his thoughts in a workshop with neighbors and colleagues.

Thus, networked relationships among tsugoes of stakeholders and of their actions is worth visualization, because it may enable to understand conflicts and interdependencies of actions with each other. Whereas conflicts may fundamentally be expressed in argumentation theories as attacks to other arguments, creative communication toward the birth of new and admissible business means to externalize hidden constraints and intentions and to agree with their importance, until the intentions and demands come to be compatible under the same constraints. For example, to Mr.X above, habitants may say the building cannot stand under the constraint that they do not like to live with such a tall building. The new action of Mr.X, as mentioned above, may come out for satisfying this demand. More smartly, Mr.X should have understood the habitants are really demanding for (i.e., intending to have more) sunshine, and try to be compatible with the condition that subshine become shed to the surrounding residential area, without failing to satisfy his own intention to expand business.

An HTN is visualized from the recorded log of actions (utterance/buying/selling) based on the following three assumptions:

- ・The action of Mr/Ms X represents X's intention, put as X_int.
- ・The pre-constraint on X, put as X_pre, should have occurred before X_int.
- ・The action on X, reflecting X_int, may be followed by X's post-constraint on other actions (of oneself or others), put as X_pst

Reflecting these assumptions, the procedure for visualizing an HTN is as follows:

Step 1) Produce a dataset where each line includes one action (utterance, when the dataset is a log of discourse) of a participant.

Step 2) Replace all lines with the intention of its actor (e.g. X) i.e., X_int.

Step 3) To each line, insert Y_pre where Y is the person who acted just after the current line. Also insert Z_pst where Z is the one who acted just before.

Step 4) To each line, insert the content of the previous line - for reflecting lingering effects.

For example, suppose Mr.X above and colleagues are talking as follows:

Ms.Z (habitant): Why don't you allow your staffs to telecommute?
Mr.X: I will take your idea. Then we do not need a big building.
Mr.U (working for Mr.X's firm): Telecommuting…? Doesn't it weaken our teamwork?
Mr.X: I will then certainly introduce telecommuting partially.
Mr.U: Partially… to what part?
Mr.X: Maybe the salesforce and the research team.
Mr.Y: Is it allowable to have them carry out the data …?
…(continues)

D =

X_pre	Z_int		Z_pre		
U_pre	X_int	Z_pst	X_pre	Z_int	
X_pre	U_int	X_pst	U_pre	X_int	Z_pst
U_pre	X_int	U_pst	X_pre	U_int	X_pst
X_pre	U_int	X_pst	U_pre	X_int	U_pst
Y_pre	X_int	U_pst	X_pre	U_int	X_pst

(4.1)

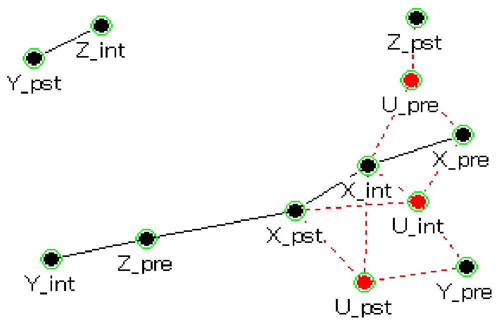


Fig. 4.1 The HTN obtained for the sample discourse

In Step 1, the second column in Eq.(4.1), framed by the bold line, is obtained. In Step 2 and 3, the first three items in each line as in the thinner frame are inserted. In Step 4, the last three items outside the frames are added for representing the lingering influence of each event on later events. By applying KeyGraph to data D, Fig.4.1 is obtained. For example, Mr.X's action may affect Ms.Z, as the link from X_pst to Z_pre shows. Also the conflict between Mr. X and Mr. U is presented by the link between X_pst and U_pst, implying X and U are influencing on each other, rather than accepting each other's requirement or solutions.

Efficiency of Innovation estimated with HTN

Here let us introduce the concept sticky information (Hippel 1994) that means the tendency of information to be localized in the brains of individual people, who may be either an inventor on the industrial (developing, producing, and selling) side or a consumer. The information about the

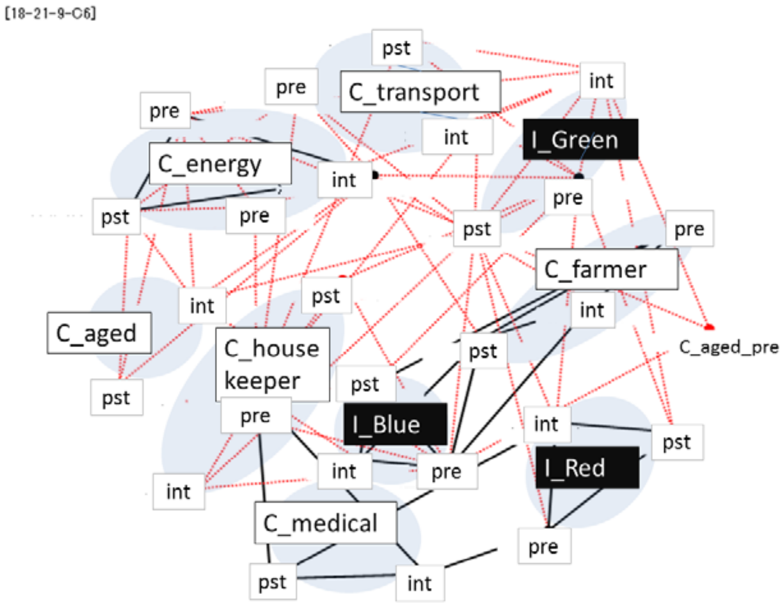


Fig.4.2 The map of KeyGraph for words and participants for one hour from the beginning of a workshop of 2 hours: Hereafter, I_xxx means inventors and C_yyy consumers. In this graph, each shadowed region shows a set of nodes representing the action of a certain stakeholder (e.g., C_farmer means the consumer who played the role of farmer, and I_red an inventor named Mr/Ms. Red) and his/her underlying intention, pre-constraint, and post-constraint.

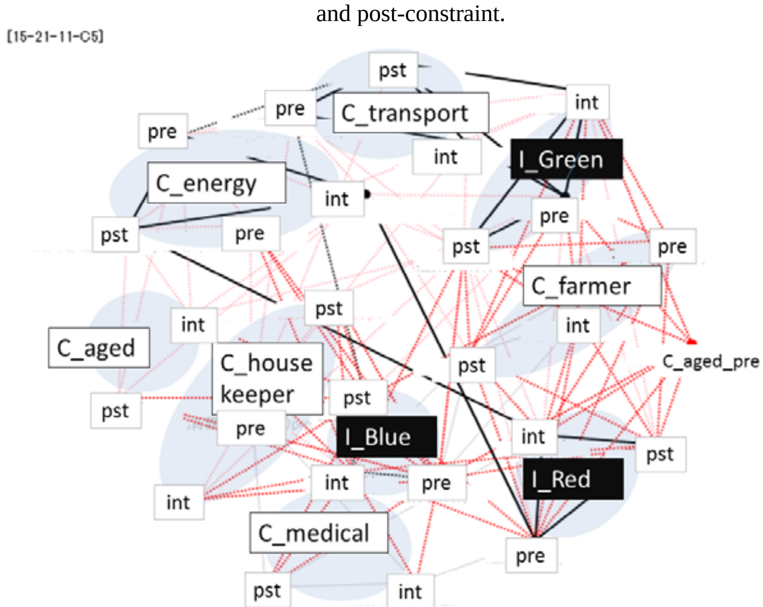


Fig.4.3 . The map of KeyGraph, for the last one hour of the same workshop as (continuing from) Fig.4.2.

requirements of consumers and the technological knowledge of inventors are hard to be transferred from/to each other, so the mutual understanding of stakeholders of each product/service tends to be disturbed. This happens even between inventors, so sometimes collaborations in a firm come to be disturbed. In successful cases, even if useful information is sticky, users may propose designs to

satisfy their own requirements, whereas manufacturers may improve products from solution-intensive viewpoint: This may be good enough if the proposal of user is easy to implement by applying established techniques and if users get satisfied by improving the efficiency. However, this does not stand if users' ideas are not feasible or practical, or if technologies of inventors are not easy to understand.

Reading cases presented in the literature (Hippel 1994, 2006, Ogawa 2010), we find the exemplified pieces of sticky information are relevant to underlying reasons for decisions and actions, similarly to tsugoes. For validating this point, Ohsawa, Horie and Akimoto (2012) visualized an HTN for the discourse log of each hour in workshops where participants discussed strategies toward innovation (Ohsawa and Nishihara 2012). In these workshops, communication was processed between "inventors" who create ideas of businesses and "consumers" who evaluate, criticize, or ask questions about the proposed ideas. As a result, the authors found the tsugoes of participants tend to be connected to each other in cases where created ideas at last come to be highly evaluated after the workshop. That is, information relevant to tsugoes should be externalized and communicated, i.e., unstuck, in order to achieve successful results of workshops. For example, the most highly evaluated idea created by inventor I_Red, who finally came to be ranked at the top, was to "reflect human's body temperature to automatic air conditioning" that was evaluated highly by consumers who talked on behalf of energy-industry, medical worker, and aged people.

The tsugoes of I_Red used to be linked with medical workers at first as in Fig.4.2, but came to be linked with aged people and energy industry workers as in Fig.4.3. I_Green, who finally came to be evaluated as the second best inventor in this workshop, used to be linked to energy industry at first in Fig.4.2. Then, he shifted to be linked more tightly with a transportation company. His best evaluated idea was "a transportation system enabling to call buses anywhere and take the most efficient route." Although this is apparently addressed to the transportation industry, this was also evaluated highly by the energy industry. In HTN, such a connection among each inventor and consumers who bought the inventor's ideas emerged in all cases where obtained ideas were evaluated relative highly than in other workshop cases. For 32 ideas obtained in 5 workshops, nearly half (51.07%) of consumers who purchased each idea were linked to its creator in the corresponding HTN, regardless of the conditions set for the included workshops. Thus Ohsawa et al showed the structure of the network of tsugoes is quite relevant to the quality of created ideas from the aspect of innovation, i.e., practical change in business.

TDC and HTN

Here let us compare between TDC and HTN (note: mere application of TDC in visualizing the time sequence of HTN, such as Fig. 4.2 and Fig. 4.3, produced no meaningful results according to our experiments).

We can say TDC (Sugimoto et al 2012) is a method to visualize the time sequence of the structural changes in event-to-event relationships, showing spots where significant events may be missed. Especially, the missed event may mean a noteworthy trigger to contextual shift. An “event” here can be replaced approximately with a “human” by regarding a human’s action as an event. For example, the utterance “Mr.X said he will build a tall building” can be abstracted into just “Mr.X did something” or only “Mr.X”. Thus, for just investigating who is affected by whom, the dummy nodes in TDC means a force of such as a hidden interest or a hidden leadership is affecting the members taking part in the conversation. On the other hand, HTNs visualize human-to-human structures with showing links corresponding to the links among tsugoes, which are hidden but significantly meaningful for humans’ decision making and actions.

Thus, we can say TDC and HTN applied to the log of participants’ behaviors/arguments in a community both visualize hidden factors potentially embracing a significant impact to the events/actions in the future. And here we should take into account that links between humans shown in an HTN show essential contacting partners to unstick information, i.e., share tsugoes, with each other for realizing innovation. In this sense, if the dummy nodes shown by TCD corresponds to hidden tsugoes, then we can expect the investigation of the real entity or real events corresponding to the red (dummy) nodes in TDC should enable to know who should talk to whom for realizing an

innovative discussion.

Furthermore, in the recent challenge we are integrating TDC with analysis and visualization of the discourse structure in arguments, taking account of attacks and defenses, data supporting the premises (prior conditions) and effects (post constraint) of each argument, as by Kubosawa et al (2013). And, here an attack is essentially the stimulus to point out misunderstood or ignored intentions and pre-/post-constraints as we can read in (Atkinson, Bench-Capon, McBurney 2006). That is, the relationships among arguments attacking each other mean partners whom one should contact and urge to speak out (externalize) tsugoes. Thus, in short, in the period of this project, the progress of tsugology outside of our project also came to support the direction of our extension in the future toward innovative communication.

5. Communication Analysis using a ubiquitous sensor

This research aims at getting the whole picture of the dynamics of internal organization activities by analyzing an organization from the bottom up based on the activity status of each individual in the organization and microscopic information regarding interactions among organization individuals by using ubiquitous sensors.

We used two types of small wearable sensor devices, an acceleration sensor and a nearby sensor, in order to measure the activities of individuals and the interactions among individuals of the organization. These sensor devices consisted of an acceleration sensor and a nearby sensor. The acceleration sensor measures the physical activity level of each individual from changes in their acceleration velocity, which resulted in obtaining a time series of their physical activity level. The nearby sensor observes the adjacency of other sensor devices using infrared technology. This makes it possible to obtain the history of face-to-face contacts between individuals. Therefore, the physical activity level obtained through the acceleration sensor indicates the activity status of each individual, while the history of face-to-face contacts obtained through the nearby sensor indicates the interaction status of individuals.

In this experiment, 136 employees work for this company and we collected data from them for 53 consecutive days (33 days except holidays). ID-card type sensor devices were used in this experiment. Each employee wore this sensor device around the neck so that it would be positioned on their front chest. One sensor device was assigned to each employee. Arriving at the office, each employee picked up their own sensor device from the battery charging booth and returned it when they left the office.

The acceleration sensor was a triaxial sensor. This sensor calculated the physical activity level by counting the zero-crossing frequency, after the value has been converted to a single axis. In this experiment, one minute was given as the unit time. This one minute was divided into six sections, where each section has 10 seconds. Of these six sections, the number of zero crossings in the 10-second section that had the largest number of zero crossings was considered to be the physical activity level of that one minute.

The infrared sensor used in this experiment detects the contact with another sensor device within the range of about 2m and toward the front. Similar to the physical activity level, in each minute, the sensor-device IDs of other employees contacted during this one minute were recorded. In this experiment, we implemented an interpolation processing so that the contact state between two individuals should be detected mutually.

A feature amount of the physical activity level
With regard to the physical activity level, the following feature amounts were focused on:

(1) Standard deviation of the nonzero value of the physical activity level
(2) Standard deviation of the duration of the active state

The active state, in above (2), is defined as follows: With the average of the nonzero values of the physical activity level of the day as the threshold, the time for the physical activity over the threshold is regarded as the active state, otherwise it is regarded as an inactive state. For instance, Figure 5.1

shows the physical activity level of a person for a day. The feature amount (1) indicates variations in amplitude of the physical activity level while the feature amount (2) indicates variations in time of the physical activity level.

In order to calculate the above-mentioned two feature amounts, we used the following two methods.

> **Method A**: To calculate the feature amount at once targeting at the physical activity level data for the entire time during the measurement period
> **Method B**: To calculate the feature amount based on the physical activity level data of the day each day in advance, and then obtain the average value later

Method A indicates the feature of quantitative changes in the physical activity level, while method B indicates the feature of changes in the activity level of a normal day.

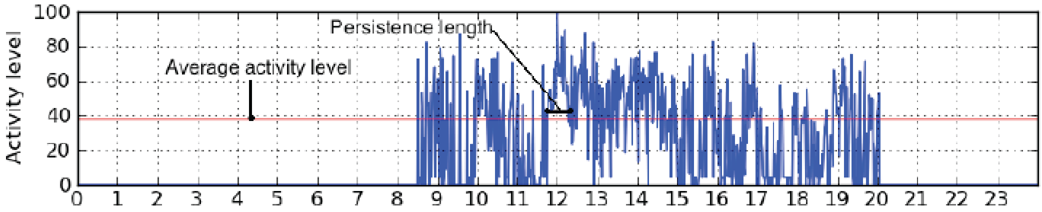


Fig. 5.1: An example of the physical activity level of a day

A feature amount of face-to-face contacts

As the feature amount of face-to-face contacts, the time for fact-to-face contact between two individuals was normalized doubly with the time that both individuals for wearing the sensor device (namely the time for being in the office), and afterwards, a total was obtained for each individual. This value was obtained based on the contact matrix as follows.
<Step1> Sum each contact matrix over the entire measurement period.
<Step2> Square each component of the matrix.
<Step3> Divide each component of the matrix with the sensor-wearing time of two individuals that correspond to each line and each row, respectively.
<Step4> Total each component of the matrix according to the line, and this value should be the feature amount of each individual that corresponds to each line.

Each component is squared in step 2 so that the value becomes dimensionless when the value is normalized doubly in step 3. This feature amount probably becomes the indicator that expresses the amount of face-to-face contacts considering differences in times of each individual for being in the office. Hereafter, this indicator is referred to as a face-to-face contact ratio.

Relationships between Physical Activity Levels and Face-to-Face Contacts

The relationships between the physical activity levels and face-to-face contacts of individuals are examined in this section. Setting 134 individuals with an adequate amount of sensor data as the sample, correlations in each feature amount between the physical activity level and face-to-face contacts, defined in the previous section, were confirmed. Table 5.1 shows the Pearson product-moment correlation coefficient between feature amounts and the significance test results. A significant coefficient with the face-to-face contact ratio was observed between the standard deviation of physical activity level (B) and the standard deviation of the duration of the active state (B). Despite

Table 5.1: Correlations in the physical activity level and face-to-face contacts

	Correlation with the face-to-face contact ratio			
Feature amount of physical activity level	Correlation coefficient	t value	df	p value
Standard deviation of physical activity level (A)	0.09	1.04	132	0.3011
Standard deviation of physical activity level (B)	0.54	7.37	132	1.657e-11

Standard deviation of the duration of the active state (A)	0.03	0.34	132	0.7308
Standard deviation of the duration of the active state (B)	0.54	7.37	132	1.657e-11

the fact that both methods calculated the standard deviation of the physical activity level and the standard deviation of the duration of the active state, a strong correlation was observed in the feature amount calculated by using method B, when compared to the feature amount calculated by method A.

These results showed us that the face-to-face contact ratio is very closely related, not to the qualitative feature of the physical activity level, but to the feature of changes in the physical activity level of a day. The time series of physical activity levels suggested the possibility of estimating the amount of communication of a certain individual with other employees in the organization.

Analysis of Face-to-Face Contacts

The history of face-to-face contacts was given to the face-to-face matrix per the unit time. The time series of the face-to-face matrix was sliced according to the predetermined time width, and each slice was treated as an adjacent matrix. We interpret a sequence of adjacent matrix as face-to-face contact history as following example.

t1 { A,B }
t1 { C,D,E } At t1, A and B had a meeting, and C, D and E had a meeting in parallel.
t2 { A,B} At t2, A and B continued their meeting.
t3 { B,C,D }
t3 { E,F } At t3, B,C and D had a meeting, and E and F had a meeting in parallel.
t4 { E,F } At t4, E and F continued their meeting.

This form is similar to the form of discussion record. As we have introduced so far, TDC is a useful tool to extract several features from the discussion record. Therefore, we thought TDC is a promising method to extract human relation from contact history.

We obtained a contact history data of a research section of a certain company. In this section, there are 17 people such as a secretary, a manager, senior researchers and junior researchers, and they work in the same big room. This section has 2 or 3 one-year projects, and each researcher belongs to one or two projects. In this section, there are several meetings such as administrative meetings, project meetings, personal meetings and so on. Most of them in this section participate in the administrative meeting, and only project members participate in the project meetings. Sometimes two projects have the joint meetings.

The size of our data is very huge because the number of contact data of one day is 144 matrices, and we have data about for 2 years. Therefore, the total number becomes about 26000 ($144\times30\times12\times2=25920$). We divided total data into 25 periods, which means each period contains one month contact data. Then, for each period, we measure the similarity between researchers by the total number of meeting in which both researchers participate. Based on the similarity value, we classify the individuals into several groups. In the case of the discussion record, each cluster corresponds to a topic which appeared in the record. On the contrary, in the face-to-face contact history, each cluster corresponds to a project team.

Fig 5.2, Fig.5.3 and Fig.5.4 are results of TDC of continuous three months. In these terms, B, K, L and P belong to the same project, and E, I, J, N and O belong to another project. We see that we can estimate the project of each researcher to some extent. However, in detail level, this method is insufficient because activities of project members are complicated. They may communicate each other not only by face-to-face contact but by emails. Nevertheless, in Fig 5.5 (December 2010), we find that the result of TDC is very different from that of August. This is because in December, a lot of small meetings are held, so clustering result becomes unstable.

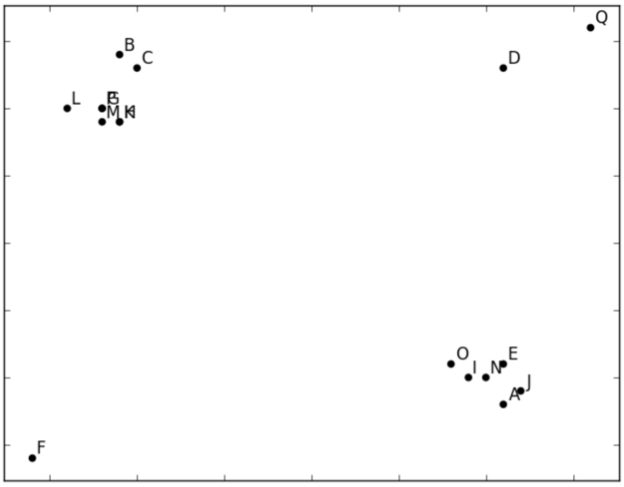


Fig. 5.2 Contact on June 2010

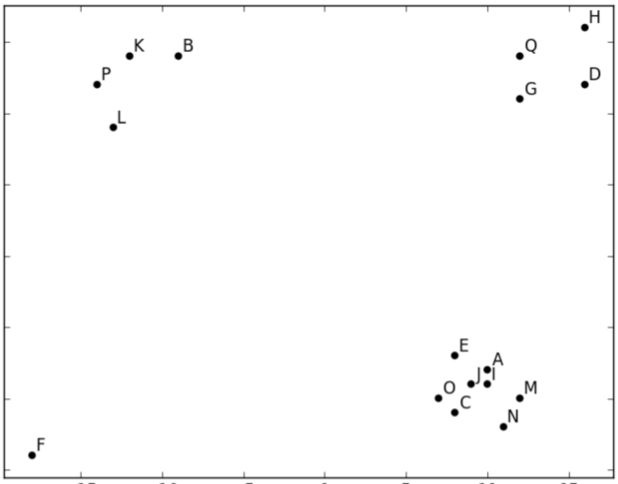


Fig.5.3 Contact on July 2010

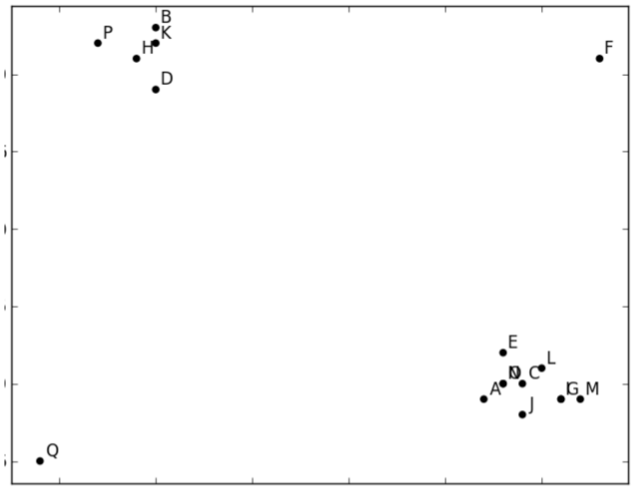


Fig. 5.4 Contact on August 2010

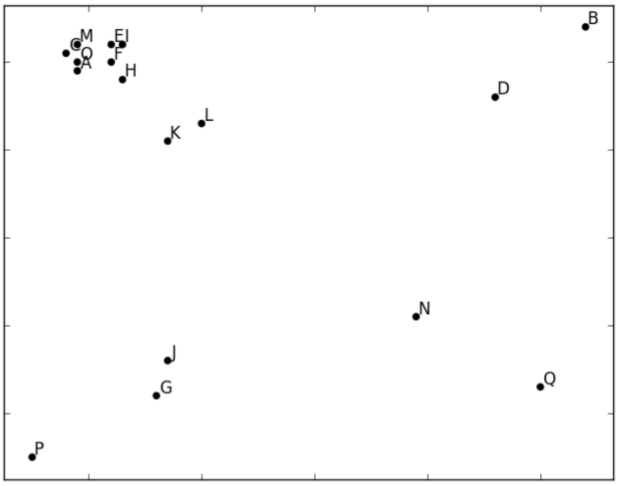


Fig.5.5 Contact on December 2010

In the case of discussion record analysis, after clustering words, in order to detect key utterances, we calculated ranking function for each utterance. Also, in the case of contact analysis, in order to detect periods when meeting pattern changed a lot, we calculate ranking function for each period as follows.

For person X (X = A,B,C,…), we calculate the meeting time with others during period *t*, *meeting(*,t)*, and makes a meeting vector *M(X,t)*.

$$M(X,t) = \begin{pmatrix} meeting(A,t) \\ meeting(B,t) \\ meeting(C,t) \\ \cdots \end{pmatrix} \qquad (5.1)$$

The ranking function *Ich(X,t)* is defined by the cosine transformation between *M(X,t)* and *M(X,t+1)* . We don't calculate the cosine transformation when |*M(X,t)*| is small because *Ich(X,t)* becomes sensitive to noise.

$$Ich(X,t) = \begin{cases} \dfrac{M(X,t)\cdot M(X,t+1)}{|M(X,t)|\cdot|M(X,t+1)|} & |\mathrm{M(X,t)}| \geqq \theta \\ -1 & |\mathrm{M(X,t)}| < \theta \end{cases} \qquad (5.2)$$

Fig.5.6 represents the value of *Ich(X,t)* (ranking value) of each month. In this figure, where the color is dark blue, the ranking value is -1. Where the color is light blue or yellow, the ranking value is low (meeting members are relatively stable). Where the color is brown and red, the ranking value is high.

As A is a secretary, her ranking value is -1 in most months. On the contrary, the ranking value of M, who is a manager, is relatively high. Other people hold meeting with similar members from April 2010 to September 2010, and meeting members are not stable from December 2010 to March 2011. And from April 2011 to July 2011 meeting members are relatively stable, but from August 2011 to March 2012, they become unstable, again. Ranking values of E, F, G, H, I, J and L are sometimes -1 because they are visiting researchers and when they are absent, there is no contact data for them.

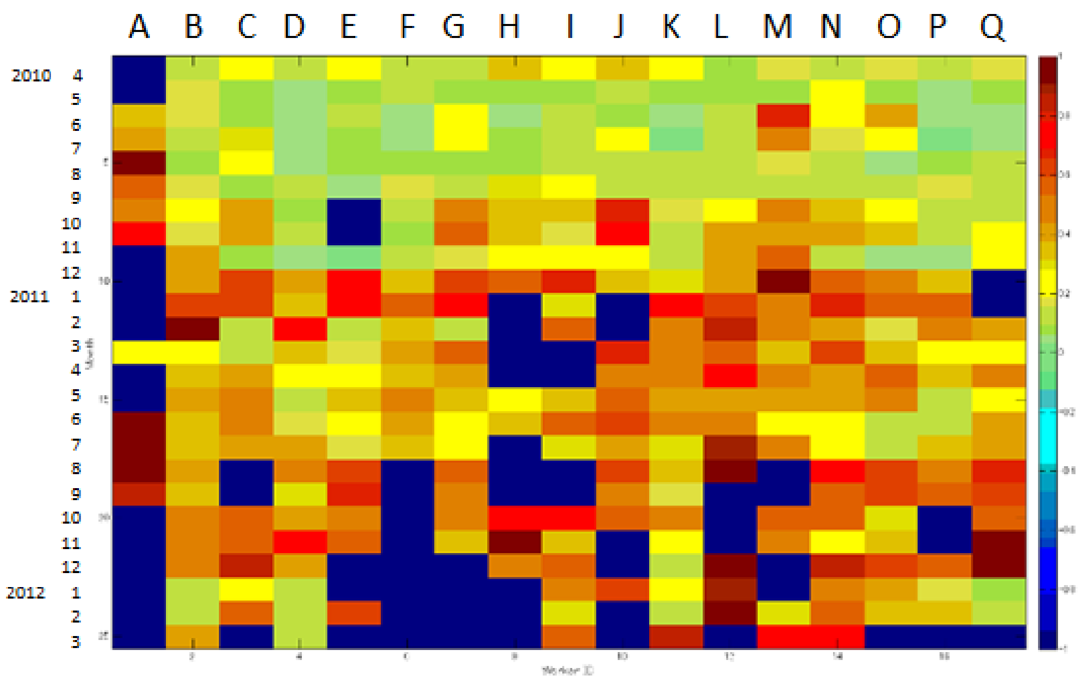


Fig. 5.6 Detecting period of group change

Results and Discussion:

In this project, to improve the preciseness of TDC analysis and to expand its application field, we studied following 4 research themes.

(1) Meta-level Discussion Analysis

Me-level TDC is a method to compare more than one discussion records. By comparing outputs of TDC method, we showed it may revive the missed dummy nodes. This method uses not only the text records but additional information such as speech acts and factors. Therefore, Meta-level TDC requires troublesome annotation task. However, in spite of such task, Meta-level TDC is promising method, because by this additional information, not only results of TDC method become more reliable, but time sequence analysis becomes possible.

(2) TDC using Multi-modal Information

Usually, when some utterances are presented with some actions, we assume those utterances are more important than others. Our method interprets the meaning of dummy nodes by considering the existence of nonverbal information. This doesn’t change the original framework of TDC. By an experiment of TV debate program, we showed this method improves the precision rate of dummy nodes. Moreover, we showed a way to detect the heat-up scene using nonverbal information. Detecting heat-up scene is useful information to manage discussions.

(3) Human Tsugo Network

Tsugology is a new method for improving and optimizing the effect and efficiency of communications and of actions in business/politics/etc based on communication, via the externalization (verbal expression and recognition) of stakeholders’ tsugoes. We analyzed human networks based on Tsugology by TDC and showed that the Human Tsugo Network which presents relation among stakeholders’ intensions and actions is effective to visualize communication for negotiation and design task. We will continue the research of Human Network of various communication tasks.

(4) Communication Analysis using a Ubiquitous Sensor

We focused on two ubiquitous sensors, an acceleration sensor and a nearby sensor. As the raw data of these sensors are huge, we developed a way to extract features from these data. And we showed there is a correlation relation between two kinds of data. Though this is the preliminary study to analyze the workers' life style in the office, the current research result is promising to find the way to improve the workers' environment in the office. Furthermore, we analyzed the history of meetings, and showed we can estimate the research group by TDC to some extent.

List of Publications:

(b) papers published in peer-reviewed conference proceedings

[1] Satoh, T., Okada, S., Nitta, K., Deliberation Process Support System for Citizen Judge Trial Based on Structure of Factors, 5th International Workshop on Juris-Informatics (Juririn 2010), Dec. 2011.

[2] Takahiro Ueda, Masaki Sugimoto, Shogo Okada, Yukio Ohsawa, Yoshiharu Maeno and Katsumi Nitta, Discussion Analysis Using Temporal Data Crystallization, 6th International Workshop on Juris-informatics (Jurisin 2012), Dec, 2012.

(c) papers published in non-peer-viewed journals and conference proceedings

[3] Sugimoto, M., Okada, S., Nitta, K. , Discussion Analysis Considering Verbal and Non-Verbal Information (in Japanese), The 26th Annual Conference of the Japanese Society for Artificial Intelligence, 1B1-R-3-2, Jul. 2012.

[4] Nitta, K., Okada, S., Sugimoto, M. , Discussion Analysis Considering Logical Structure and Emotion (in Japanese), SIG-SLUD, The Japanese Society for Artificial Intelligence, Mar. 2012.

[5] Nitta, K., Multimodal Discussion Analysis Based on Temporal Sequence, in *Advances in Chance Discovery* (ed. Ohsawa, Y. and Abe, T.) (ISBN 978-642-30113-1), Springer Verlag, Jul. 2012.

References

Shumpei Kubosawa, Youwei Lu, Shogo Okada and Katsumi Nitta, Argument Analysis with Factor Annotation Tool, The 25th International Conference on Legal Knowledge and Information Systems (JURIX 2012), 2012.

Youwei Lu, Shogo Okada and Katsumi Nitta, Semi-supervised Latent Dirichlet Allocation for Multi-label Text Classification, The 26th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE) 2013.

Shumpei Kubosawa, Kei Nishina, Masaki Sugimoto, Shogo Okada and Katsumi Nitta, A Discussion Training Support System and Its Evaluation, The 14th International Conference on AI and Law (ICAIL 2013),

Shogo Okada, Yusaku Sato, Yuki Kamiya, Keiji Yamada, Katsumi Nitta, Analysis of the Correlation between the Regularity of Work Behavior and Stress Indices Based on Longitudinal Behavioral Data, 14th ACM International Conference on Multimodal Interaction (ICMI) 2012

Takanori Sato, Syogo Okada and Katsumi Nitta , " Deliberation Process Support System for Citizen Judge Trial Based on Structure of Factors" , Proc. 5th International Workshop on Juris-informatics (JURISIN 2011),(1 Dec, 2011)

Atkinson, K., Bench-capon, T., McBurney, P., Computational representation of practical argument, *Synthese*, vol.152, pp.157–206 (2006)

Carrol, JM. (2000), Making Use:Scenario-based design of Human-computer Interactions, The MIT press

Goldratt, EM. (1987), Essays on the Theory of Constraints, North River Press

Kushiro, N., and Ohsawa, Y. (2006), A scenario acquisition method with multi-dimensional hearing and hierarchical accommodation process, New Mathematics and Natural Computation, 2 (1), pp.101-113

Nonaka, I., A dynamic theory of organizational knowledge creation, Organization Sci. 5(1), 14-37 (1994)

Ohsawa, Y., and Akimoto, M., Unstick Tsugoes for Innovative Interaction of Market Stakeholders, *International Journal of Knowledge and Systems Science*, Vol. 4, Issue 1, pp. 32-49 (2013)

Ogawa, S.. (2010), Does Sticky Information Affect the Locus of Innovation? Evidence from the Japanese Convenience-Store Industry, *Research Policy* 26(7-8), 777-790 (1998)

Ohsawa,Y. (2005), Data Crystallization: Chance Discovery Extended for Dealing with Unobservable Events, *New Mathematics and Natural Computation* 1(3), .373 - 392

Ohsawa, Y., Horie, K., and Akimoto,, M. (2012), Sticky Tsugoes underlying Sticky Information, In *Proc. 16th International KES Conference on*

Ohsawa, Y., Nishihara,Y., Nakamura,J., and Kushiro, N., Nitta, K. (2010), Tsugology for Revealing Intentions and Constraints, in *Proc. IEEE Conf. Systems, Man, and Cybernetics (SMC)* 1332-1337

Ohsawa, Y., and Nishihara, Y. (2012), *Innovators' Marketplace: Using Games to Activate and Train Innovators* (Understanding Innovation), Springer

Polanyi,M. (1966, 2009), *The Tacit Dimension*. London, Routledge. (University of Chicago Press, 1966, 2009 reprint)

von Hippel, E. (1994). "Sticky Information" and the Locus of Problem Solving: Implications for Innovation. *Management Science*, 40(4), 429-439

von Hippel, E. (2006), “Democratizing Innovation” The MIT Press; New Ed

Attachments: Publications listed above.
Copies of papers [1] – [5] are attached. As [3] and [4] are published in Japanese, they are translated into English.

DD882: Attached to this report

Deliberation Process Support System for Citizen Judge Trial Based on Structure of Factors

Takanori Sato[1], Shogo Okada[1], and Katsumi Nitta[1]

Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology, Japan
{tsato,okada,nitta}@ntt.dis.titech.ac.jp

Abstract. In 2009, the Japanese government adopted the citizen judge system. In this system, three professional judges and six citizen judges listen to the arguments between the prosecutor and the attorney, and decide the judgment through the discussion in deliberation. However the presiding judges have not been trained for moderation sufficiently, so that their skills of moderation affect the performance and quality of the discussion. Therefore, in this paper, we propose a deliberation process support system. The system assists the presiding judge to facilitate the deliberation by some functions. First it visualizes an argument structure graph, representing summary of the arguments between the prosecutor and the attorney. Next it recommends topics which need argument and their order. We propose a novel algorithm to select these topics based on preliminary research. Finally it also recommends speakers who did not have the opportunity to make remarks. Here we introduce this system.

Keywords: Argument Visualization, Argumentation, Moderator Support

1 Introduction

In May 2009, the Japanese government adopted the citizen judge system. Citizen judges chosen from ordinary people started to participate in trials as judges. At first in the trial, the prosecutor argues. Next, the attorney counters the prosecutor's argument. After that, professional judges and citizen judges have a discussion based on the arguments. It is called deliberation. In deliberation, they decide whether the accused is guilty or innocence. If guilty, they also decide the punishment. The presiding judge plays a key role as the moderator.

In deliberation, several problems exist. At first, many topics are intricately related to other ones [1]. Therefore, the citizen judges are confused about the topics under discussion on deliberation because they deal with the huge quantity of information. Second, the presiding judge needs to moderate the discussion that many individuals participate in [2]. The presiding judges have not been trained for moderating discussion, so that their skills of moderation affect the performance and quality of the discussion. Next, the time for deliberation is limited. Therefore, the presiding judge needs to insure that the discussion is effective during the limited time. Finally, the citizen judges have very little knowledge of

the law. Therefore the presiding judge needs to inform them of what the legal knowledge is required.

To solve the above problems, it is necessary for the presiding judges to select topics properly. In addition, the presiding judge has to give a fair chance to remark and proper advice to each participant. In the situation described above, a system to support the presiding judge in deliberation is needed.

For developing this kind of the system, it is promising that it has functions of argument visualization and moderator navigation. In related studies, Reed et al. proposed Araucaria [3]. This system analyzes arguments and visualizes them as a diagram. It is used for education and intended for use as argument analysis. In addition, Loukis et al. proposed 'Computer Supported Argument Visualization' (CSAV) [4]. It is a system that was for remote support of legislation debate. It has focused on argument visualization. Nohara et al. proposed an method of argument using the "chart method" in deliberation [1]. The "chart method" is a series of methods, making "a chart" and processing the deliberation using it. This approach allows the participants to share information, so that they can grasp which topic is discussed. Hotta quantitatively analyzes simulated deliberations in the citizen judge system [2]. He uses the quantity of utterances in some simulated deliberations and compares between the characteristics of the participants and it. In addition, Anzai et al. developed an annotation system for the citizen judge system. It can annotate records of the deliberation and visualize information about the deliberation [5]. This system is for analysis as to what is good deliberation. As mentioned above, there have been extensive researches done regarding argument analysis and visualization. However systems with the ability to make navigation to moderator are rare.

In this paper, we propose a deliberation process support system for the presiding judge to carry out the deliberation smoothly. The system visualizes the argument summary as a graph. In addition, it gives some recommended information to the presiding judge to moderate discussion smoothly.

In section 2, we introduce the system outline. In section 3, we show the factor registration editor. In section 4, we show the deliberation process support system. In section 5, we give our conclusion.

2 Overview of Proposed System

As our deliberation process support system uses a factor based approach, we define a factor at first. Then, we show the overview of our system.

2.1 Factor

A factor is a proposition representing a fact, an opinion, a topic, or a claim. It has additional information. We define it on the bases of the factor in [6].

A factor has information such as "ID", "state", "meaning", "type", "support", and "conflict". "ID" is the ID to identify each factor. "State" refers to the position of the person claiming an issue, taking "k", "b", or "o" as the public prosecutor's claims, the attorney's claims, or the other state. "Meaning" is a description of the factor. "Type" refers to the type of the factor and is registered as one of the following three types.

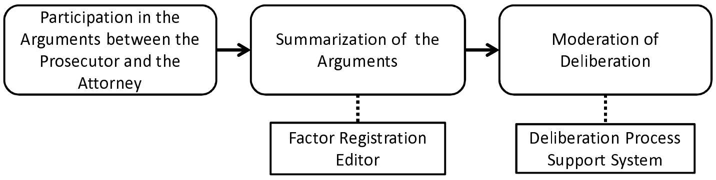


Fig. 1. Step Flow of the Citizen Judge Trial Using the Deliberation Process Support System

- Penalty factor
  This factor is related to the legal issues. It is what the prosecutor and the attorney claim as to what punishment is appropriate.
  e.g. "The accused is to be in prison in five years.", "The accused should be declared innocent."
- Main factor
  This factor is needed to think whether the accused is guilty or not, or whether the punishment is serious or not. It assists penalty factors.
  e.g. "The accused had a motivation.", "Average punishment of similar incidents is 3 years."
- Evidence factor
  This factor represents the fact, the testimony of witnesses, the evidence of the case, and so on. It assists main factors.
  e.g. "The fingerprints of the accused were found on the knife left at the crime scene."

"Support" refers to support factors, which current factor assists as the evidence or the cause. "Conflict" refers to conflict factors, with which the current factor conflicts.

2.2 Overview of Citizen Judge Trial Using Deliberation Process Support System

We propose a deliberation process support system. The target user is the presiding judge. To use the system, there are three steps. Fig.1 shows the step flow.

At first, the judges participate in the arguments between the prosecutor and the attorney. Next, the presiding judge uses a factor registration editor to summarize the arguments. Then, the system makes a factor list, a collection of factors. After creating the factor list, the editor outputs an argument structure graph. In deliberation, the presiding judge moderates the deliberation using the graph. He/she inputs factors which appeared in the participant's remark in remark record table. The system visualizes old remarks on the table. In addition, it analyzes the graph and the table to notify the user of two types of recommendation. One recommendation is the factors which need arguments and their order. The other is the speakers who have not had the opportunity to speak about each factor. The user decides the next factor or speaker with the information and facilitates the deliberation. After the participant make a remark, the phase of inputting the remarks is again used and the cycle repeats.

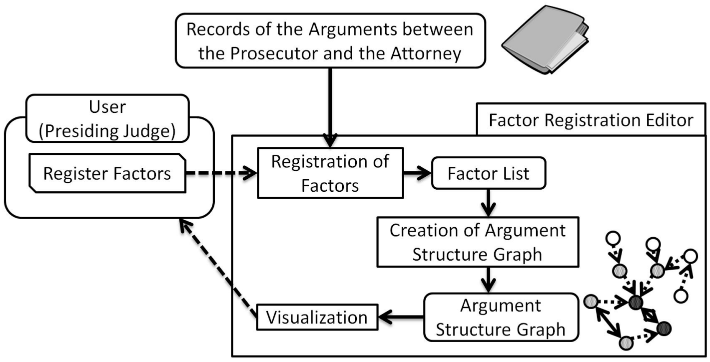


Fig. 2. System Architecture of the Factor Registration Editor

The editor and the system were developed using Java and prefuse [7][1].

3 Factor Registration Editor

3.1 Overview

This editor helps the user to register factor to summarize the arguments between the prosecutor and the attorney. The editor has the following functions.

- Registration of factors and creation of argument structure graph
  The user registers factors based on records of the arguments between the prosecutor and the attorney. A collection of the registered factors is called a factor list. The system makes a graph structure, called a argument structure, from the factor list.
- Visualization of argument structure graph
  It visualizes the argument structure graph, representing the relationship of the factors based on the factor list.

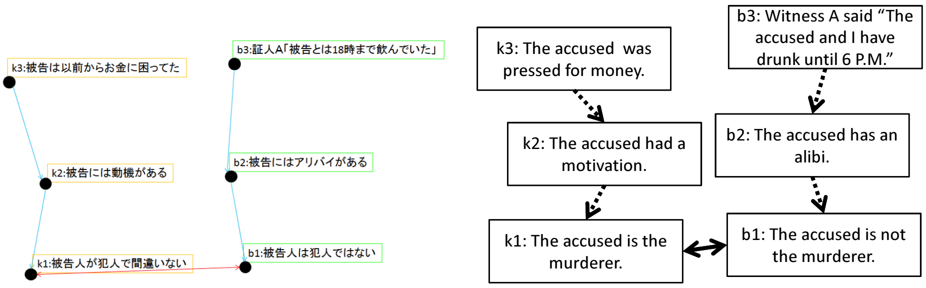
Fig.2 shows system architecture of the editor. The editor displays records of the arguments and helps the user register factors. Then it creates argument structure. In registering factors, it displays the graph. After the user registers the factors, the editor outputs the graph data.

The editor displays records of the arguments as text, factor configuration space, and factor list. In this space, the user can register or modify factors and the factor list, which is a collection of registered factors.

3.2 Registration of Factors and Creation of Argument Structure Graph

The user inputs records of the arguments to the editor. The user registers factors by referring to the records. Specifically to register factors, the user inputs information such as “ID”, “state”, “meaning”, “type”, “support”, and “conflict” described in section 2.1. The registered factors are included in factor list.

[1] For more information and download, access to `http://prefuse.org/`



(a) Original View (b) Translation in English of (a)

Fig. 3. The Argument Structure Graph Displayed in the Factor Registration Editor

After the user finishes registering factors, the system makes the factor list into a graph. This graph is called the argument structure graph. Fig.3 shows a sample of the graph visualized by the editor. It consists of nodes and edges. Each node represents a factor. It has information, node ID, which contains "state" and "ID", and "meaning". On the other hand, each edge represents a relationship between factors. It has information of the type of relationship. The edge registered "support" represents a directed dotted arrow. In addition, the edge registered "conflict" represents a bidirected solid arrow. If looking at the visual graph when creating factor list, the user can consider the type of relationship between a factor being registered and factors that have already been registered.

The argument structure graph can be thought to summarize the arguments between the prosecutor and the attorney. The user can use the graph to proceed with arguments in the deliberation later. To do this, the editor has the function to output data to be input to the deliberation process support system. This data is represented by the GraphML. It has information required to produce the graph.

4 Deliberation Process Support System

4.1 Overview

The system supports the user to facilitate moderation during deliberation. The system functions are as follows.

- Visualization of argument structure graph
  It gets a file of argument structure graph which is output from the factor registration editor and visualizes the graph.
- Support of the input of participants' remarks
  It provides an table in which the user inputs factors which appeared in the participants' remarks.
- Next topic recommendation
  It indicates which factors should be discussed in deliberation.
- Next speaker recommendation
  It indicates who should make remarks.

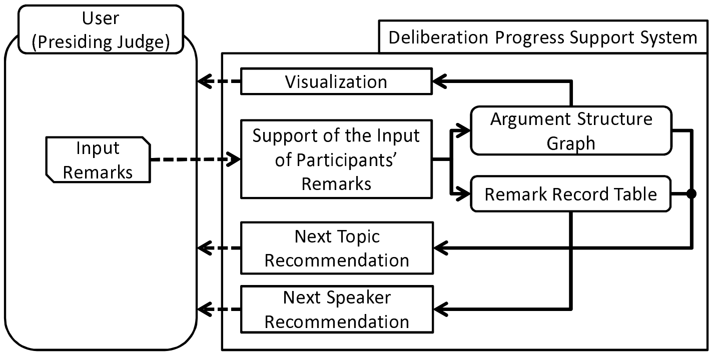


Fig. 4. System Architecture of the Deliberation Process Support System

Fig.4 shows the system architecture. Before deliberation, the system gets the data of the argument structure graph and displays the graph. In addition, the user sets the deliberation limit time as remaining time. During deliberation, the user inputs the participants' remarks in a table, called a remark record table. Using some data, the system decides factors to be discussed and speakers required to make remarks. After that, it notifies the user of the factors and the speakers. It displays the argument structure graph and the menu to input various operations.

4.2 Visualization of Argument Structure Graph

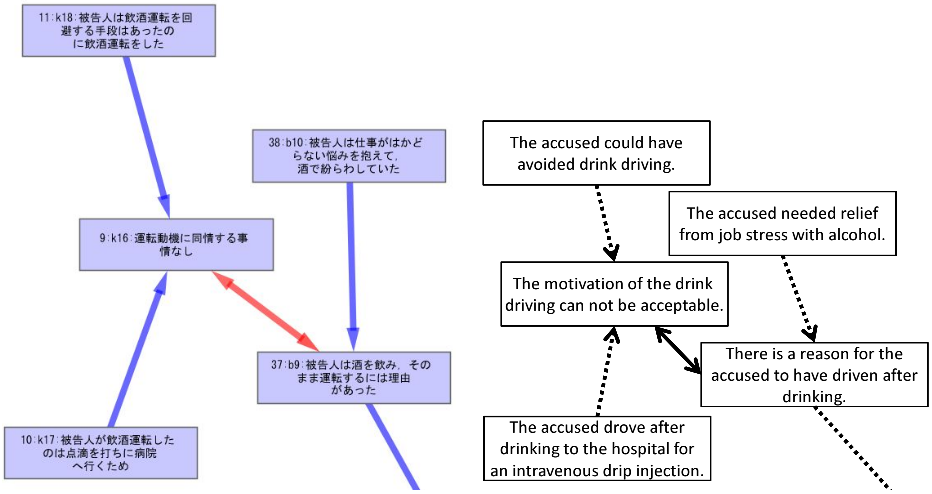
The system receives the data created from the factor registration editor and displays the graph.

Fig.5 shows a diagram of the graph when the data is input in the system. The input data is created from moot court. The number of factors is 49. As well as the factor registration editor, the system allows the user to add or remove a factor on the menu during deliberation. Furthermore, if it is difficult to see the graph, it is possible to control it for easier viewing. For example, when factors overlap because the edges are too short, the user can solve the problem by lengthening the edges. The user can moderate the deliberation by confirming which factor is being discussed by using this function.

4.3 Support of the Input of Participants' Remarks

In deliberation, the presiding judge listens to participants' opinions and summarizes them. The summarized opinions are basis of adjudication. This is called fact-finding. In Japan, presiding judges are required to do so by the law. Hence presiding judges have a heavy burden from this work. So the system offers a record table for summarizing opinions.

The user inputs the remarks or the opinions of factors in the remark record table. The table consists of rows of the factors and columns of the participants. The user fills in each participant's remarks or opinions about each factor on a corresponding cell. More specifically, the user clicks the node from the graph and fills in a displayed cell where the participant's remark meets the corresponding



(a) Original View (b) Translation in English of (a)

Fig. 5. Argument Structure Graph Displayed in Deliberation Process Support System

factor. When confirming remarks of the factor, the user clicks the factor, and then the system displays the remark record table of the clicked factor.

Each cell of the table has an "approval" or a "denial" tag. Each tag represents the standpoint that the participant takes. An "approval" tag represents that the participant supports the factor. A "denial" tag represents that the participant opposes the factor. "approval" cells are painted blue, "denial" cells are painted red. Furthermore, cells which are not filled in are colored in yellow, and cells which are filled in with something are colored white. The user can visually recognize the conditions of the cells.

4.4 Next Topic Recommendation

The system has a function to recommend factors which are to be discussed. Thus, the user complies with the recommended factors so as to moderate deliberation smoothly.

The recommended factor is selected by the factor selection algorithm. This algorithm uses the argument structure graph. It outputs recommended factors and their order.

Preliminary Research When making the algorithm, during preliminary research, we examined the difference of factor selection between a professional of the law and an amateur in the moot court.

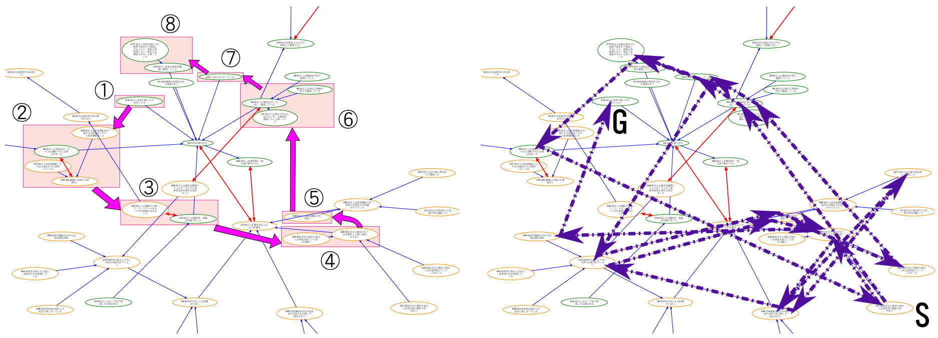
At first, we collected the simulated deliberation records. The one moderated by a law professional acquired from the records of a moot court of citizen judge system. The records included arguments between the prosecutor and the attorney, and the deliberation. In order to compare the deliberation record, we held simulated deliberation moderated by an amateur and got the records. In the real deliberation of citizen judge trial, three professional judges (including a

presiding judge) and six citizen judges participated. However, in the simulated deliberation moderated by an amateur, only a presiding judge and two citizen judges participated. The role of the presiding judge was to moderate the deliberation and to argue his/her opinions. The role of the citizen judge was to only argue his/her opinions. They summarized their opinions and decided which the accused was guilty or not and the punishment if guilty. A flow of the simulated deliberation by amateurs is described below. At first, we explained the summary of citizen judge system, the basic mechanism of the citizen judge trial, and the basic method of deliberation to the participants. Then, they read records of the arguments between the prosecutor and the attorney. The records were parts of the moot court mentioned above. Next they discussed the simulated case and decided whether the accused was guilty or not. If guilty, they continued the discussion and decided the punishment. We recorded the discussion using a microphone and a video camera. In the deliberation done by amateurs, nine subjects (graduate students, age 22 to 24 years, eight males and one female) participated. Then they were divided into three groups and a record was taken from the deliberation of each group.

After collecting the deliberation records, we analyzed the difference between deliberations moderated by the law professionals and by the amateurs. For the analysis we created the argument structure graph from the simulated case. In the graph there are 48 factors. The graph has factors about punishment such as "it is appropriate for the accused to be in prison for eight years" and "the claim that the punishment is too severe", abstract factors such as "the victim's wife dose not heal the pain of her mind" and "the accused deeply regrets the situation occurred and has apologized", and evidence factors such as "the accused drank much and drove" and "the private settlement was finished". Based on the graph, we checked the transitions of the arguments developed in the collected deliberation.

Fig.6(a) shows the transition moderated by the law professional on the graph. Encircled numbers on Fig.6(a) shows the order of the transition. As to the deliberation moderated by the law professional, the discussion proceeded by dealing with multiple factors as a group, called a factor group, and arriving at an opinion by each factor group, so that factor groups represent boxes in Fig.6(a). The arrow indicates the transition of the factors on deliberation. Besides Fig.6(b) shows the transition of the deliberation moderated by the amateur. The deliberation record is on behalf of the three deliberation records. Additionally in Fig.6(b), "S" and "G" indicate the first and last factor in the transition.

As to the deliberation moderated by the law professional, the moderator confirmed factors to be discussed before the deliberation, so that he/she could show the factors one by one and listened to the participants' opinions to summarize the factors showed. The selected factors were comparatively abstract factors supporting the factors about the punishment. On the other hand, in the deliberation moderated by the amateur, the participants did not concentrate on one factor but expressed their opinions about the punishment, so that the transition was very unsettled. Furthermore selection regularity of factors especially



(a) Factor Transition in Deliberation Moderated by a Law Professional

(b) Factor Transition on Deliberation Moderated by Amateur

Fig. 6. Difference of Factor Transitions

was not found. From the above, it is important for factor selection to choose comparatively abstract factors and to summarize the factors one by one.

Factor Selection Algorithm This algorithm operates based on the following hypotheses and the result of the preliminary research. First, factors related to many conflict factors are topical in the arguments between the prosecutor and the attorney, so that they are to be discussed first. Among them, factors related to the many support factors are the major and core topics of the case, so that they are to be discussed early. On the other hand, with regard to factors not related to conflict factors, the factors which are supported by fewer factors have less impact on the graph, so that they are easier to handle in the deliberation. Therefore they are picked up early.

Factor selection algorithm is described as below.

- Initial Input
  Graph $G = (V, E)$ consisting of a set of factor node $V \ni v_i, v_i = (nid_i, sug_i, abst_i)$ and a set of edge $E \ni e_j, e_j = (eid_j, vs_j, ve_j, rel_j)$, where we define nid_i, sug_i, $abst_i$, eid_j, vs_j, ve_j, and rel_j. nid_i is the node ID and equal to i. sug_i is the state of the factor, taking "*prose*", "*attor*", and "*other*" as the public prosecutor, the attorney, and the other. $abst_i$ is the level of abstraction, taking "*punish*", "*main*", and "*support*" as penalty factor, main factor, and evidence factor. eid_j is the edge ID and equal to j. vs_j is the head factor node ID of the edge. ve_j is the tail factor node ID of the edge. rel_j is the relationship between vs_j and ve_j, taking "s" and "c" as support and conflict.
- Algorithm
  1. $V_m = \{v_i \in V | \exists i, abst_i = \text{"}main\text{"}\}$ is taken.
  2. The parameters $f_s(v_i)$, $f_a(v_i)$ from each $v_i \in V_m$ are calculated.
     - Overall support factor number $f_s(v_i)$
       $f_s(v_i)$ is the number of the counted factors supporting the intended

factor, including $f_s(v_i)$ of the factors supporting the intended factor. It is defined by (1).

$$f_s(v_i) = \begin{cases} 0 & (|V_{v_i}| = 0) \\ 1 + \sum_{v_j \in V_{v_i}} f_s(v_j) & (|V_{v_i}| \geq 1) \end{cases} \tag{1}$$

$$V_{v_i} = \{v_i | \exists i, j, ve_j = v_i, rel_j = \text{``}s\text{''}\} \tag{2}$$

- Conflict factor number $f_a(v_i)$
  $f_a(v_i)$ is the number of the counted factors conflicting with the intended factor. It is defined by (3).

$$f_a(v_i) = |\{e_j | \exists i, j, rel_j = \text{``}c\text{''}, sug_{vs_j} \neq sug_{ve_j}\}| \tag{3}$$

Furthermore, the element $rank_i$ is added to each $v_i \in V_m$, so that $v_i = (nid_i, sug_i, abst_i, rank_i)$. Each $rank_i$ will be substituted for either number $1, 2, ..., |V_m|$, which represents argument order.

3. $1, 2, ..., |V_m|$ is substituted ascending for $rank_i$ of v_i which $f_a(v_i) \geq 1$ in order of larger $f_a(v_i)$.
   - If $f_a(v_i) = f_a(v_j)$ $(i \neq j)$,
     The factor of lesser f_s in v_i and v_j is selected and substituted first.
     * If $f_s(v_i) = f_s(v_j)$ $(i \neq j)$,
       The order of v_i and v_j is decided randomly.
4. $1, 2, ..., |V_m|$ is substituted ascending for $rank$ of v_i which $f_a(v_i) = 0$ in order of the lesser $f_s(v_i)$.
   - If $f_s(v_i) = f_s(v_j)$ $(i \neq j)$,
     The order of v_i and v_j is decided randomly.

– Output
  V_m consisting of v_i having $rank_i$

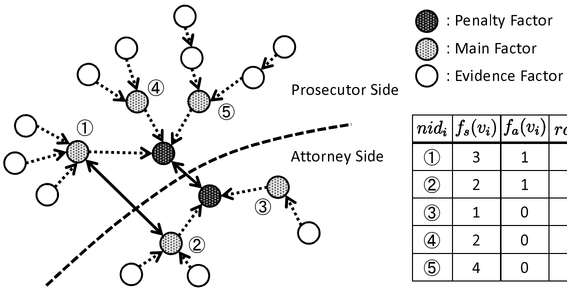
After factors are ordered in the algorithm, time constraint information is also added to them. It is named discussion time. Discussion time $t(v_i)$ is defined by (4) and assigned to each factor.

$$t(v_i) = \frac{T_s}{|V_m|} \tag{4}$$

In (4), T_s is the remaining time, the whole time that is available to be spent on the deliberation. The user inputs this time.

Finally the system recommends $v_i \in V_m$ in order by $rank$. Moreover, if the discussion time of each factor is more than $t(v_i)$, the system notifies the user that the recommendation moves on to the next factor.

Fig.7 shows an example of factor selection algorithm operation using an example of the argument structure graph. The table in Fig.7 represents parameters of main factors ①, ②, ③, ④, and ⑤. First, the main factors are taken. Second, the two parameters $f_s(v_i)$, $f_a(v_i)$ for each of the taken factors are calculated. The result of the value is shown in the table in Fig.7. When considered visually,



nid_i	$f_s(v_i)$	$f_a(v_i)$	$rank_i$
①	3	1	1
②	2	1	2
③	1	0	3
④	2	0	4
⑤	4	0	5

Fig. 7. Example of Factor Selection Algorithm Operation

$f_s(v_i)$ is the number of all subsequent nodes that support v_i, including indirect support nodes. On the other hand, $f_a(v_i)$ is the number of nodes that conflict with v_i. After the calculation of the parameters, ① and ② is selected because their $f_a(v_i)$ are larger. Among them, ① is selected because its $f_s(v_i)$ is larger, and 1 is substituted for its $rank_i$. Next, 2 is substituted for ②'s $rank_i$. Then, there are factors ③, ④, and ⑤ which do not relate to conflicting factors. Among them, their $rank_i$ are arranged in ascending order of the lesser $f_s(v_i)$. Hence 1, 2, and 3 are substituted for $rank_i$ of factors ③, ④, and ⑤.

4.5 Next Speaker Recommendation

In deliberation, it occurs that the discussion proceeds while the participants can not make remarks about his/her opinions. This situation should be avoided as much as possible. The presiding judge should give a chance to listen to the opinion of the appropriate participants in this situation. The deliberation process support system has a function to make recommendations to give an opportunity to make remarks to the participants who were previously unable to give their opinions. This function uses a speaker selection algorithm to select the participants. The system confirms the status of remark record table and makes a recommendation to give opportunity to make their comments to those participants whose table is not filled in.

The speaker selection algorithm is described below.

- Initial Input
  Remark record table $R \in r_{i,j}$, where $r_{i,j}$ is a remark record cell of the participant j related to a factor i.
- Algorithm
  1. The user inputs the notification to the system the end of argument about factor i'.
  2. $R' = \{r_{i,j} \in R | i = i', r_{i,j} = \text{“”}\}$ is taken, where “$r_{i,j} = \text{“”}$” represents that $r_{i,j}$ has no remark record.
- Output
  The system lets the user confirm the remarks by the participants j corresponding to $r_{i,j} \in R'$.

By giving proper opportunities to make remarks in this way, it is possible to have the discussion while listening to not only the participants who make many remarks, but also the ones who do not.

5 Conclusion

In this paper, we proposed a deliberation process support system for the presiding judge to facilitate discussion in the deliberations of citizen judge trial. The system uses an argument structure graph based on the arguments between the prosecutor and the attorney. Then the presiding judge uses a factor registration editor. The editor supports the user to register factors and their relations from records of the arguments between the prosecutor and the attorney. After that, it creates a factor list and the argument structure graph. The deliberation process support system uses the graph and supports the user. The system has some functions, such as visualizing the graph, supporting the input of remarks by participants, recommending the topics to be discussed and the participants that are required to make remarks.

Acknowledgment

The authors thank Prof. S. Hotta from Meiji University for his generous support.

References

1. Nohara, K., Morimoto, I., Mishima, S., Takeuchi, K.: Expansion of Deliberation Design Theory in the Citizen Judge System (Final 3.) Towards Realization of Deliberation Based on Closing Argument Analysis - Concrete Methods and Challenges and Prospects in Designing Deliberation (in Japanese). Horitsujiho 81(10), 84–95 (2009)
2. Hotta, S.: Linguistic Theories in Legal Settings (in Japanese). Hituzi Syobo (2010)
3. Reed, C., Rowe, G.: Araucaria: Software for argument analysis, diagramming and representation. International Journal on Artificial Intelligence Tools 13(4), 961–979 (2004)
4. Loukis, E., Xenakis, A., Tseperli, N.: Using Argument Visualization to Enhance e-Participation in the Legislation Formation Process. In: Macintosh, A., Tambouris, E. (eds.) Electronic Participation, Lecture Notes in Computer Science, vol. 5694, pp. 125–138. Springer Berlin / Heidelberg (2009)
5. Aanzai, K., Nitta, K.: Analyzing system of conference record of citizen judge trial based on annotation. In: Proceedings of the SIG-KBS conference. vol. 90, pp. 31–36. Japanese Society for Artificial Intelligence (2010)
6. Aleven, V., Ashley, K.D.: Teaching CaseBased Argumentation Through a Model and Examples: Empirical Evaluation of an Intelligent Learning Environment. In: Proceedings of AIED 97 World Conference. pp. 87–94. IOS Press (1997)
7. Heer, J., Card, S.K., Landay, J.A.: prefuse: A Toolkit for Interactive Information Visualization. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 421–430. CHI '05, ACM, New York, NY, USA (2005)

Discussion Analysis Using Temporal Data Crystallization

Takahiro UEDA [1], Masaki SUGIMOTO [1], Shogo OKADA [1], Yukio OHSAWA [2], Yoshiharu MAENO [3] and Katsumi NITTA [1]

[1] Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology
[2] University of Tokyo,
[3] Artificial Intelligence Group

{ueda, sugimoto, okada, nitta}@ntt.dis.titech.ac.jp
ohsawa@sys.t.u-tokyo.ac.jp

Abstract. This paper introduces a discussion analysis tool which extracts topic flow and important utterances from a discussion record based on word occurrences. We have already proposed a discussion analysis method called Temporal Data Crystallization (TDC). This method divides the entire discussion record hierarchically at points where the topic changes, and analyzes some features of flow of topics for each period. In this paper, we showed the effect of hierarchical division by analyzing an example discussion record. Then, we introduced the extension of TDC by considering nonverbal information such as actions, facial expression, loudness of voice, and so on.

Keywords. Topic Flow, KeyGraph, Data Crystallization, discussion analysis, word clustering

1 Introduction

Recently alternative disputation resolution (ADR) has been becoming popular instead of judicial trials. Especially, mediation is one of the promising methods to build consensus. However, the analysis of mediation records is more difficult than that of arbitration records. In the case of arbitration, it may be sufficient that we observe only the logical aspects of arbitration records because participants don't need to reach the consensus. However, in the case of mediation, we have to observe not only the logical aspects, but emotional aspects of the discussion records.

Though there are several tools to analyze discussion records, most of them treat mainly the logical aspects based on Toulmin Diagram [1]. Therefore, we started the research for developing tools to analyze emotional aspect of mediation records. The basic technology of our research is temporal data crystallization (TDC) [4] which is an extension of the data crystallization method (DC) [2][3]. These methods have been devised to observe hidden intention from discussion records based on co-occurrence of words. These methods extract candidates of key utterances from the mediation records in the form of dummy nodes. The key utterances include ones which change topics, ones which refer to another topic, ones which the speaker put stress on, and so on.

The main problem of DC method is it sometimes extracts non key utterances as dummy nodes. At first, we show that TDC method can extract more correct key utterances than DC method by analyzing an example discussion records. Then, we show it is possible to extract more correct key utterances by using nonverbal information.

In Section Two, the key concept of TDC is introduced. In Section Three, the usefulness of TDC by comparing coherence of word clusters of TDC and original DC is evaluated. In Section Four, to improve the precision of clustering words which appeared in the record, TDC is expanded by considering multi modal data.

2 Analysis on Co-Occurrence of Words

2.1 Word Clustering Analysis

Ohsawa and Maeno have proposed a method of analysis on points to be discussed by utilizing a word clustering method with data crystallization [4]. By doing so, they showed that these methods can be utilized for extracting the speaker's hidden intensions. Based on this word clustering method, we tried to enhance this method in order to improve the extraction accuracy of hidden intensions and to extract important utterances.

Word Clustering with Data Crystallization (DC) is performed as follows. Where the discussion record is considered to be a set of S_1, S_2…, and each utterance S_i is considered to be a set of words that appeared, $\{w_1, w_2…,w_n\}$, the method proposed by Maeno et al. defines the distance $d(w_i,w_j)$ between each word as the reciprocal of the Jaccard coefficient. Next, all words that appeared in utterances are clustered into the given number $|C|$ $(C_1, C_2,…,C_{|C|})$, by utilizing the K-medoids method (Fig. 1). When each word is expressed with a node and words having a high Jaccard coefficient are connected with links, a graph that consists of n islands (clusters) can be obtained. Each cluster is probably considered to be a single topic.

Next, for each utterance $S_i(i=1,2,…)$, following ranking functions is calculated. Here, $c(Si)$ is the number of words belonging to Si, and ε is an constant.

$$\mathrm{I}_{av}(S_i) = \frac{1}{|C|}\sum_{j=0}^{|c|-1} C\left(S_i \cap c_j\right) \quad (1)$$

Formula (1) is used to find an utterance S_i which contains multiple clusters inside. We select some utterances whose ranking value are relatively high, and for each selected utterance S_k, we insert a dummy node d_k in the graph.

The appearance of these dummy nodes suggests that the utterance that corresponds to these nodes refers to several topics. This indicates that other topics are mentioned during the utterance about a certain topic, or a topic is guided to transition to another topic.

The use of dummy nodes provides the possibility of discovering the characteristics that are not expressed on the surface of the utterance record. For example, Maeno has shown the possibility of extracting the hidden intentions contained in a utterance by utilizing dummy nodes. This is because topics that attract

attention and interest can be predicted by making utterances that contain related words even without making clear utterances.

Fig.2 shows an example of a word clustering graph with dummy nodes.

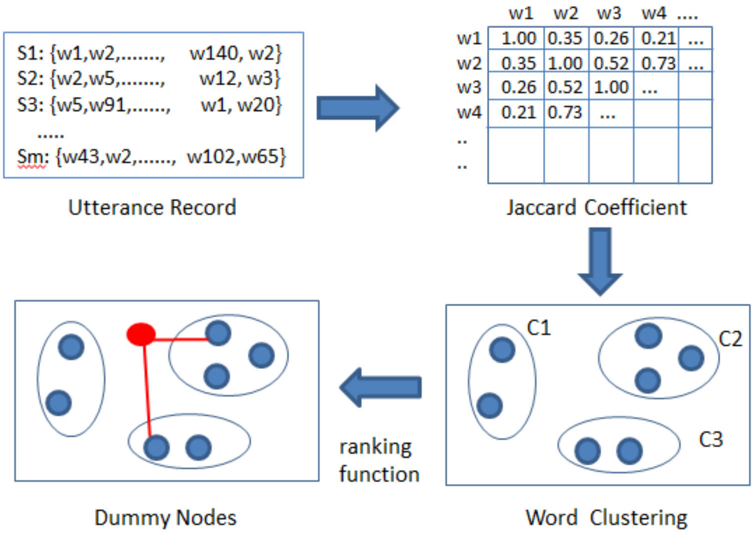


Figure 1: Word clustering and dummy nodes

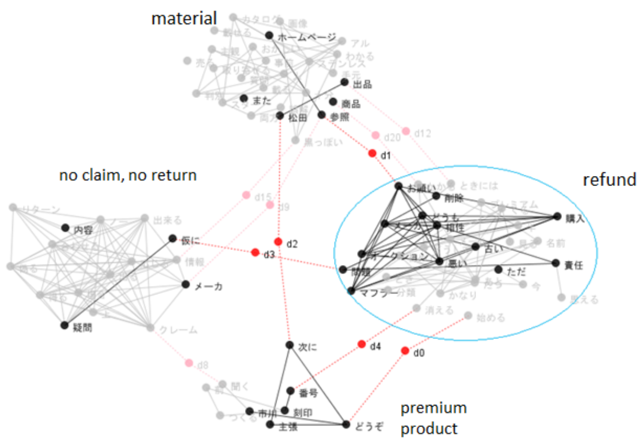


Figure 2: Example of Word Clustering and Dummy Nodes

2.2 Temporal Word Clustering (Time-series Word Clustering) with Temporal Data Crystallization (TDC)

Word clustering shown in the previous section is an effective method for analysis of topic transitions based on the discussion record. However, this method has the issue

that the clustering precision might decrease when the discussion extends for a long period of time and contains a lot of topics, along with complicated topic transitions. For example, depending on the words, it could be natural to classify words into different clusters between the first half and the last half of the discussion. However, the above-mentioned clustering method can only classify words into the same cluster through the entire discussion.

Given this issue, we proposed two ways to enable this word clustering method to handle the passage of time. One method is, when the Jaccard coefficient is calculated, to consider not only co-occurrence within the same utterance, but also co-occurrence between adjacent utterances. The other method is to divide the discussion record at each point where topics make a significant shift in order to re-cluster words according to each individually divided section.

In this research, we proposed the latter method (Temporal Word Clustering with Temporal Data Crystallization (TDC)) [2]. This method is performed as follows. At first, by applying the word clustering method with DC for the entire discussion record, the words that appear are divided into a given number of clusters (Fig.3). Next, a histogram, which shows how words appeared in each cluster as time passed, is obtained. This histogram shown as bar charts indicates each of the clusters. When there is a point where two lines clearly cross, this point is determined to be where topics made a significant shift. Before and after this point, the discussion record is divided into two sections, and then the word clustering method is applied to each of these divided sections. Afterwards, repeating this process divides the discussion record in a hierarchical way.

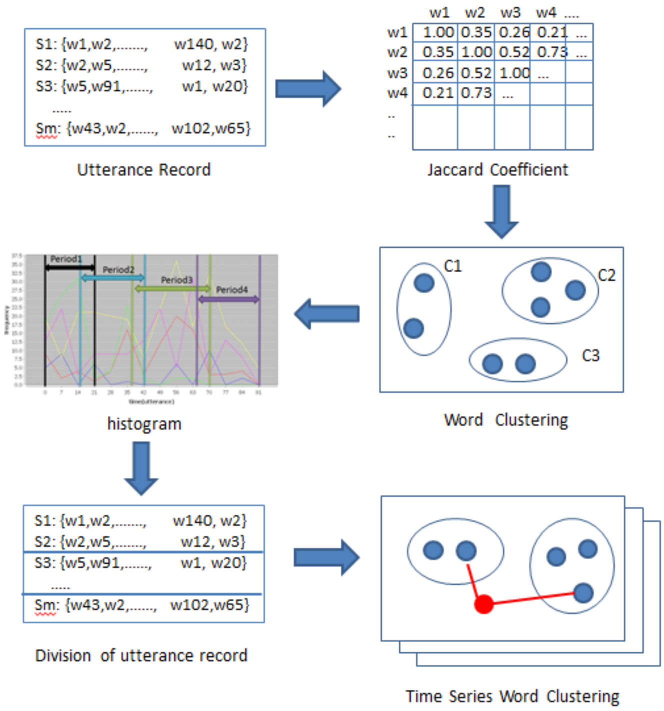


Figure 3: Time series word clustering

Just as described above, those topics that are discussed partly within each section can be made obvious by analyzing each individual section. In addition, dummy nodes that make strides over both sections can be extracted, by overlapping a few utterances existing around the section borders of both sections when sections are divided. Compared to those dummy nodes existing within the sections, those dummy nodes making strides over multiple sections can be considered to be utterances that are related to the shift of topics.

3 Evaluation of Temporal Word Clustering

In this section, we show the effects of temporal word clustering by applying it to discussion records of a mock mediation conducted by law school students. The trouble case treated by the mock mediation is as follows.

Mr. X put an automobile muffler on the auction Web site. Usually, automobile mufflers are made of stainless. However, Mr. X's muffler is made of a poor material, but Mr.X didn't show any explanation about the material. He just showed the URL about the manufacturer of the muffler. On the Web page, there is a catalogue of mufflers. But, the muffler is not listed on the current catalogue because the muffler was made as a custom made product a few years ago.

Mr. Y purchased the muffler, but he left the muffler untouched for two months. Two month later, he started using it and found that it is inferior product. He asked Mr. X to cancel the contract and return the money.

We applied the original word clustering method and a temporal word clustering method to the above mentioned discussion record and extracted some dummy nodes. We manually extracted some key utterances where the topic change occurred. We evaluated how many dummy nodes correspond to a key utterance. Table 1 shows the comparison between the time-series method and the original one.

Table 1. Effect of Time-series Method

Number of Dummy Nodes	TDC	DC
10	5/14	4/14
15	9/14	5/14

In this table, for example, when we extracted 10 dummy nodes by the time-series word clustering method, among 14 key utterances, 5 utterances corresponded to dummy nodes. According to this experiment, we could extract more precise dummy nodes by the time-series word clustering method.

4 TDC by Considering Multi-Modal Data (Discussion analysis using Multi-modal information)

In the previous section, we showed the possibility of discovering the topic transition using dummy nodes. However, analysis by dummy nodes has two problems. First problem is that dummy nodes are often affected by noise. And, the second problem is that information gained by dummy nodes is not sufficient, because information of dummy nodes is just there are utterances which refer to two topics. If we interpret dummy nodes using verbal information such as grammatical information or nonverbal information such as actions of speakers, we may extract more precise and more detail information from dummy nodes.

Recently the Japanese government has been considering the storage of the video recordings of trials, and broadcasting companies often broadcast TV discussion programs. If we use these discussion records in the form of a movie, our dummy nodes based analysis will be improved.

4.1 Extraction of Topic Transition using Gesture Information

Our target record is obtained from a discussion where each participant sits in a chair. From discussion video records, we observed salient characteristic of speaker (Table 2) and labeled each utterance (Table 3).

We show a method for extracting topic transitions. The label of gesture information $\{a_1, a_2, \dots, a_n\}$ is attribute of the dummy word d_i (formula 2).

$$S_i=\{w_1, w_2, \dots, d_i\} \quad (2)$$

First , we rank the each utterance using the ranking function(see formula 4). Let b_i be discussion records, c_j be the discussion topics, $C(b_i)$ be the number of words belonging to utterance S_i, $|c|$ be the number of topics, and ε be a constant. It then is

Table 2. Gesture Labels

Body part	Label	Meaning of the label
Head	Downward	Looking down
	Forward	Putting the head forward
	Nodding	Nodding the head
Trunk	Rightward	Tilting the trunk to the right
	Backward	Tilting the trunk backward
	Leftward	Tilting the trunk to the left
	trunk Forward	Tilting the trunk forward
	Back and forth	Tilting the trunk back and forth
Hands and arms	Hands horizontal	Moving the hands horizontally
	Hands vertical	Moving the hands vertically
	Folding	Folding the arms
	Bringing together	Bringing hands together
Voice	Loud	Speaking with a loud voice

Table 3. Gesture of Discussants

Speaker	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Downward	0	0	0	0	0	0	2	1	0	0	0	0	4	1
Forward	3	1	0	2	9	2	1	1	0	5	0	0	1	0
Nodding	0	0	0	0	1	1	6	2	0	0	0	1	0	1
Rightward	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Backward	0	0	0	0	0	0	0	0	0	2	0	0	1	4
Leftward	0	0	0	0	0	0	0	0	0	1	0	1	0	0
trunk Forward	9	1	4	1	1	1	2	4	0	24	1	0	4	3
Back and forth	1	0	0	0	0	1	0	0	0	0	1	0	0	0
Hands horizontal	0	0	0	2	0	0	3	0	0	0	0	2	1	0
Hands vertical	0	0	0	2	0	3	13	10	3	19	0	3	4	5
Folding	3	2	0	7	0	0	0	0	0	0	0	0	1	3
Bringing together	9	0	0	18	0	0	2	0	0	0	2	2	0	0
Loud	1	0	0	0	0	0	1	0	0	17	0	0	0	0
Total	26	4	4	33	11	8	30	18	3	68	4	9	16	17

possible for the utterance of top rank to guide the discussion toward transition to another topic.

$$\mathrm{I}_{av}(S_i) = \sqrt{\sum_{j=0}^{|c|-1} \left(\frac{C(S_i \cap c_j)+\varepsilon}{C(S_i)+\varepsilon|c|} - \frac{C(S_{i-1} \cap c_j)+\varepsilon}{C(S_{i-1})+\varepsilon|c|} \right)^2} \qquad (3)$$

We compared the existing method (word clustering using language only) with the proposed method (word clustering using language and non-verbal) to examine the accuracy of topic transition. In this experiment, we continued to use the discussion record that was used in the previous section (see Table 3). Table4 shows the experimental results. Formula 4 and formula 5 show calculation method 1 and calculation method 2.

$$\begin{cases} \text{The number of correct answers} = \text{Number of the topic transitions in the top N and labeled gestures} \\ \text{Precision} = \dfrac{The\ number\ of\ correct\ answers}{Number\ of\ the\ gesture\ lavel\ in\ the\ top\ N} \\ \text{Recall} = \dfrac{The\ number\ of\ correct\ answers}{Number\ of\ the\ gesture\ label\ in\ the\ topic\ transition} \end{cases} \qquad (4)$$

$$
\begin{cases}
\text{The number of correct answers} = \text{Number of the topic transitions in the top N of labeled gestures} \\[1ex]
\text{Precision} = \dfrac{The\ number\ of\ correct\ answers}{N} \\[2ex]
\text{Recall} = \dfrac{The\ number\ of\ correct\ answers}{Number\ of\ the\ gesture\ label\ in\ the\ topic\ transition}
\end{cases}
\tag{5}
$$

In case of the calculation method 1, the precision showed using our proposed method is better than the existing method. However, the recall showed the opposite results, because some utterances without gestures cannot be detected. And, in case of calculation method 2, the proposed method shows better results in both precision and recall than existing methods.

We verified that improved extraction accuracy of topic transition using not only text data, but also non-verbal information could be achieved.

Table 4. Detection of Topic Change

		Existing method	Proposing method	
			Calculation method 1	Calculation method 2
Rank 20	the number of correct answers	3	1	6
	Precision	0.150	0.500	0.300
	Recall	0.077	0.026	0.500
Rank 40	the number of correct answers	9	2	12
	Precision	0.225	0.333	0.300
	Recall	0.231	0.051	1.000
Rank 60	the number of correct answers	14	5	
	Precision	0.233	0.385	
	Recall	0.359	0.128	
Rank 80	the number of correct answers	19	7	
	Precision	0.238	0.318	
	Recall	0.487	0.180	

4.2 Extraction of Topic Transition using Speaker Pairs Information

Next, we focused on the change of *speaker pairs*. A *speaker pair* is defined as two persons speaking alternately.

Figure 4 shows that relation between a change speaker pairs and topic transition. The discussion record is the same as in the previous section. This result shows a topic transition of 45% of the total was seen when the speaker pair changed. And, all topic transitions were within six utterances of the change of speaker pair in case of this discussion record. Therefore we targeted the utterances within the six utterances of the change of speaker pair for the discovering topic transitions.

We verified that improved extraction accuracy of topic transition using speaker pair information could be achieved (Figure 5).

In this section, we showed that improved extraction accuracy of topic transition using not only text data, but also non-verbal information was achieved (Gesture and Speaker pair).

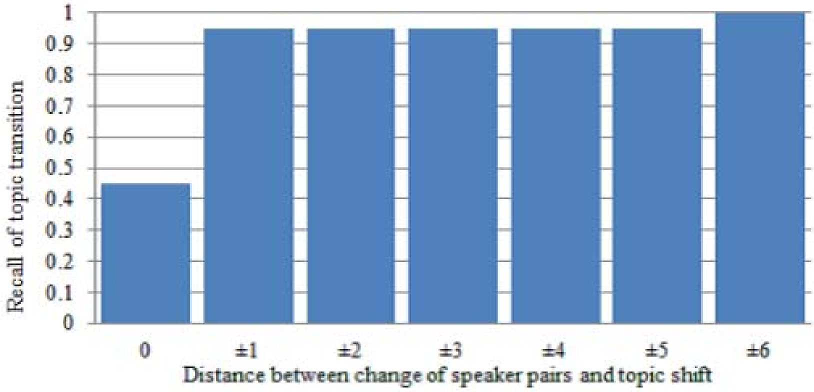


Figure 4. Relation between speaker and Topic Change

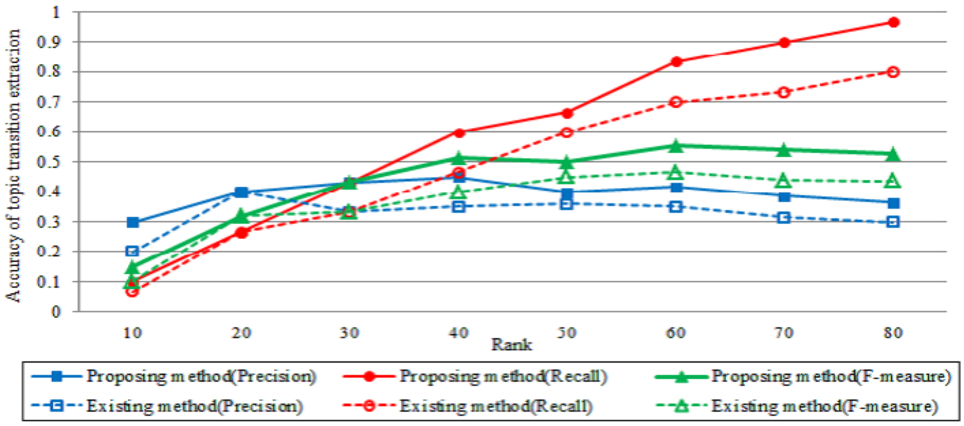


Figure 5. Comparison of Precision and Recall

5 Conclusions

As a method for analyzing emotional aspects of mediation records, we introduced the temporal word clustering with time-series data crystallization method. This method extracts topics in the form of cluster of words, and key utterances in the form of dummy nodes. By experiments, we showed the temporal word clustering method generates more correct dummy nodes than the original word clustering method. And we showed that we can generates more proper dummy nodes by considering nonverbal information such as gesture of speakers or relation between speakers and topics.

References

[1] Araucaria, http://www.computing.dundee.ac.uk/staff/creed/araucaria/

[2] Ohsawa eds. Chance Discovery in Real World Decision Making, Springer Verlag, 2006.

[3] Maeno, Y. et al., Crystallization highlighting hidden leaders, Proc. IPMU, 2006.

[4] Nitta, K. et.al, Scenario Extraction using Word Clustering and Data Crystallization, Proc. Jurisin 2009

Discussion Analysis Considering Verbal and Non-Verbal Information

Masaki Sugimoto, Shogo Okada, Katsumi Nitta

Department of Computational Intelligence and Systems Science

1. Introduction

In order to analyze discussion conducted during negotiations or arbitrations, it is important to learn the relationships and discussion structure in the topics discussed. However, when the discussion extends for long hours, it is difficult to understand the outline by reading a statement record. Additionally, learning transitions of topics and extracting those statements that triggered a transition in topics are an overburdened task to conduct. It is very difficult to learn the relationships between topics and understand the discussion content accurately, especially when the discussion contains multiple topics. There are other factors that bring about difficulty in understanding the discussion, such as an obscure statement which contains a hidden real intention.

Given these indications, there exists KeyGraph [Ohsawa 06] as one of the studies that analyzes discussion structures. In KeyGraph, those words that co-occur frequently form clusters, resulting in the expression of one topic. By utilizing data crystallization [Ohsawa 05], items that are not contained in the data could be inserted as a dummy item, which makes it possible to observe the transitions of topics as dummy nodes. However, setting parameters became a burdensome task when it was attempted to correspond clusters to topics. With the purpose of reducing this burden, Maeno proposed a topic estimation method [Maeno 07] by incorporating clustering. By using this method, words that appear could be clustered based on the number of clusters that were configured according to the estimated number of words. This word clustering made it possible to extract clusters according to the number of topics. In addition, Nitta made a study of enhancing the accuracy of topic extraction by separating discussion record based on the consideration of time series [Nitta 12].

However, these existing methods still have low accuracy in topic extraction, and a great deal of noise was contained in the dummy nodes that are indexes for topic transitions. For these reasons, manual interpretation becomes necessary, which puts a great burden on human operators. One of the reasons for this issue is that the existing methods focus only on text analysis, ignoring non-verbal information such as facial expressions and gestures of the speakers. Due to this issue, the existing methods cannot consider circumstances where statements are made, and as a result, detailed analysis cannot be conducted. In addition, according to McNeill et al., non-verbal information possibly contributes to grading and organizing verbal information. This report makes suggestions as to the importance of non-verbal information [McNeill 01].

We propose a new discussion analysis method that considers not only verbal information, but

also non-verbal information. To be specific, our proposed method focuses on co-occurrence between words that appear in statements and non-verbal information. The ultimate goal of this research is to reduce the burden on human operators for the analysis of discussion structures by adding non-verbal information, and further to enhance accuracy of the discussion analysis. During this research, we performed experiments with regard to detection accuracy of topic transitions that became necessary for understanding discussion structures. Through the experiments performed, we have confirmed that use not only of verbal information, but also of non-verbal information contributes to enhancing detection accuracy.

2. Word Clustering Method Considering Time Series

This section gives an explanation about the word clustering method considering time series that was used in our research. First, this method separates words that appeared into a given number of clusters (topics). Next, a histogram is obtained in order to indicate how words in each cluster appeared as time progresses. In this histogram, bar charts that indicate each cluster shown. If there is a point where two lines clearly and distinctly cross, it can be determined that the topic significantly changed around this crossing point, and the discussion record is divided into two sections, before and after this point. To each section that has been divided, this word clustering method is applied. Repeating this process subsequently divides the discussion record in a hierarchical way. By separating statement record at a point where the topic made a major shift and analyzing each separated section, the specific topic that is discussed within the section becomes obvious.

3. Discussion Analysis Method Considering even Non-Verbal Information

3.1 Proposed Method

Here we propose a discussion analysis method introducing non-verbal information. Specifically speaking, this method focuses on co-occurrence between words that appeared in statements and non-verbal information. In this research, we performed experiments with regard to detection accuracy of topic transitions that become necessary for understanding the discussion structures.

Dummy nodes are highly likely to appear when topics transit. Dummy nodes indicate where a certain statement contains multiple clusters, the representative word for each of the clusters that are connected together. However, dummy nodes are easily affected by noise because they are used to determine whether one statement contains multiple clusters or not. Additionally, it is impossible to discern what kind of multiple topics are included.

Recently, however, the video recording of court trials has been considered, mock trials conducted in the departments of law at universities have been recorded on video, and debate programs have been broadcast on TV. Amid such circumstances, not only text information, but also non-verbal information has been extracting from discussion records. In this research, an attempt was

made to enhance the accuracy of the appearance of dummy nodes by utilizing non-verbal information for this word clustering method.

For this research we targeted discussion data; therefore, speakers made most of their statements while sitting. For this reason, the types of gestures were limited. We focused empirically on three parts of the upper body, the head, the trunk, and the arms. From video data, we observed whether some characteristic movements could be seen when people made statements during discussion. Afterward, we manually gave 13 gesture labels to each statement. Table 1 shows 13 labels actually used for this research. The gestures considered were actually made by those who made statements; gestures of those other than who made statements were not considered.

Table 1: Gesture Labels

Body part	Label	Meaning of the label
Head	Downward	Looking down
	Forward	Putting the head forward
	Nodding	Nodding the head
Trunk	Rightward	Tilting the trunk to the right
	Backward	Tilting the trunk backward
	Leftward	Tilting the trunk to the left
	Forward	Tilting the trunk forward
	Back and forth	Tilting the trunk back and forth
Hands and arms	Hands horizontal	Moving the hands horizontally
	Hands vertical	Moving the hands vertically
	Folding	Folding the arms
	Bringing together	Bringing hands together
Voice	Loud	Speaking with a loud voice

This section describes how dummy nodes appear. Setting discussion records as a set of {S1, S2, …, Sn}, and each statement Si is identified as a set of words that appeared, {wi1, wi2,…, wim}. At that time, where labels that indicate gesture information, {a1, a2,…, ak} are the attributes of the dummy word, di, is expressed as the equation (1) below.

$$Si=\{wi1,\ wi2,\ \ldots,\ di\} \qquad (1)$$

First, each statement is ranked by using the ranking function [Maeno 07]. From statements that have large ranking of values from the number of dummy nodes specified by the user are then determined as statements in which dummy words are inserted. Next, in the statements in which dummy words are inserted, the top two words that show a strong relationship between topics are

selected. These two words are words that exist within the discussion and are linked with the dummy words. These dummy words actually appear on the graph as dummy nodes. These dummy words indicate the judgmental standards, and this means that dummy nodes do not appear unless a gesture label was given to the statement. Performing preliminary experiments, we have confirmed that in statements made accompanied by gestures, topic transitions easily occur. Originally, we considered the possibility of including non-verbal information that indicates a habit. However, since no meaningful non-verbal information related to the discussion has been specified, all gestures were counted in this research.

The next section describes the detection of topic transition as an example of analysis by using this method.

3.2 Analysis Example

As an example of our analysis, we used the record of statements from a debate TV show, “Asamade Nama TV” (A Live Telecast until the Morning). This TV show featured 14 participants including the program presenters. The topics discussed were the relocation of the U.S. air base in Okinawa, economic stimulus measures, and the consumption tax. Table 2 shows gestures (including whether the participants raised their voices) that were observed in the show. There were 1183 statements made, to which 271 gesture labels were given.

Figure 1 shows the clustering result of the first half of the show. Black nodes indicate words, while red links indicate dummy nodes. In this figure, the following topics were shown: the domestic relocation of the air base (Fig. 1: Upper left), the overseas relocation of the air base (Fig. 1: Upper right), about the Prime Minister’s Office (Fig. 1: Left), willingness of the local residents (Fig. 1: Right), about Prime Minister Hatoyama (Fig. 1: Lower left), and about the deterrent force (Fig. 1: Lower right). Here, Figure 2 shows a magnified graph centered on dummy nodes d138 and d140. The attribute “Arms Vertical” (swinging of the arms vertically) was given to statements ID138 and ID140. When the statement content 3 was seen, in statement ID138, the topic transitioned from the relocation of the air base to the willingness of the local residents. In statement 140, the topic transitioned from willingness of the local residents to the relocation of the air base. This is an example where topic transitions could be detected by giving gesture labels.

The next section examines whether the accuracy of detecting topic transitions can be enhanced by utilization not only verbal information, but also non-verbal information.

Table 2: Gestures of discussion participants

Speaker	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Downward	0	0	0	0	0	0	2	1	0	0	0	0	4	1
Forward	3	1	0	2	9	2	1	1	0	5	0	0	1	0
Nodding	0	0	0	0	1	1	6	2	0	0	0	1	0	1
Rightward	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Backward	0	0	0	0	0	0	0	0	0	2	0	0	1	4
Leftward	0	0	0	0	0	0	0	0	0	1	0	1	0	0
Forward	9	1	4	1	1	1	2	4	0	24	1	0	4	3
Back and forth	1	0	0	0	0	1	0	0	0	0	1	0	0	0
Hands horizontal	0	0	0	2	0	0	3	0	0	0	0	2	1	0
Hands vertical	0	0	0	2	0	3	13	10	3	19	0	3	4	5
Folding	3	2	0	7	0	0	0	0	0	0	0	0	1	3
Bringing together	9	0	0	18	0	0	2	0	0	0	2	2	0	0
Loud	1	0	0	0	0	0	1	0	0	17	0	0	0	0
Total	26	4	4	33	11	8	30	18	3	68	4	9	16	17

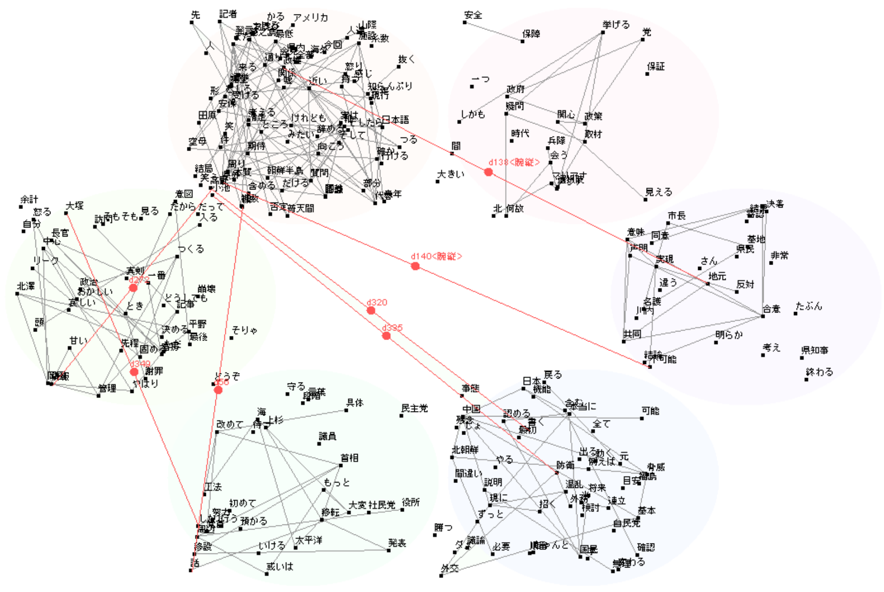


Figure 1: Word Clustering Graph of TV Debate Program

Upper left: Domestic relocation of the U.S. air base

Upper right: Overseas relocation of the U.S. air base

Left: About the Prime Minister’s Office, Right: Willingness of the local residents

Lower left: Prime Minister Hatoyama, Lower right: The deterrent force

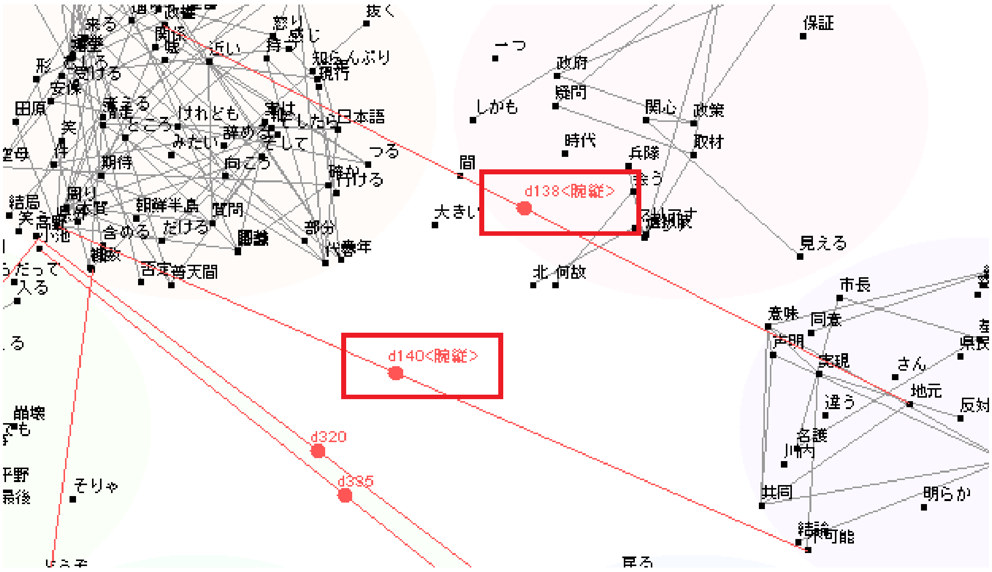


Figure 2: A magnified graph centered on dummy nodes d138 and d140

Table 3: Statements (Statements ID136 – ID141)

ID	Speaker	Content
136	Kawauchi	(Omitted) I'm saying that it is impossible to realize this plan. <Crossing his arms>
137	Tawara	This is a bit difficult to understand, we need interpretation. Um, Mr. Otsuka, what is he saying?
138	Otsuka	(Omitted) No consensus has been reached with the local residents, so it means there is no guarantee yet that the scenario goes just according to what was claimed in today's joint declaration. <Arms vertical>
139	Tawara	No, not at all.
140	Mogi	(Omitted) So, the Prime Minister said that, right? Saying what is unrealizable, he also said that at least the air base would be relocated outside Okinawa, during the election campaign. After all, this relocation was impossible. Now he said the base would go to Henoko. It's too late to refer to another destination like Henoko, it's totally impossible to relocate the base there. (Omitted) <Arms vertical>
141	Yamagiwa	(Omitted) It does not necessarily mean that all are opposed to the presence of the base. Not all. (Omitted)

4. Comparison Experiments

In order to investigate the detection accuracy of topic transition, we performed comparison experiments using the existing method and the proposed method. The existing method is the time-series word clustering method that utilizes only verbal information. In these experiments too,

we used the statement record from the same debate TV show, “Asamade Nama TV,” that was used in the previous section. Those scenes where topics transitioned were extracted manually, and considered as the correct data for topic transitions. Themes significantly changed in the first half and the last half of the statement record, and analysis was conducted on each individual section. The next section describes the results of these comparison experiments.

4.1 Comparison Experiment 1

We used the first half of the TV show in this comparison experiment 1. Here, those issues about the relocation of the U.S. base in Okinawa and Prime Minister were discussed. Table 4 shows the parameters set. The analysis targets were 349 statements, from discussion ID44 to discussion ID 392.

The introductory part of the discussion (Discussions ID0 – ID43) was omitted from analysis, because it contained the description of this debate show and self introductions of the participants. This discussion contained the following six topics: the domestic relocation of the U.S. air base in Okinawa, the overseas relocation of the U.S. air base in Okinawa, willingness of the local residents, the deterrent force, statements made by Prime Minister Hatoyama, and the Prime Minister’s Office. Table 4 shows the core words of this discussion. Core words mean possible words that are assumed in a particular topic. By determining these words preliminarily, the accuracy of topic extraction was enhanced and 150 gesture labels were given to statements. Additionally, topics transitioned 39 times. Figure 3 shows the clustering results.

Table 5 shows the results of the detection of topic transition in comparison to experiment 1. “Top 20” indicates the top 20 dummy nodes when statements, where dummy nodes were inserted by using the ranking function and sought out. In the existing method, the number of correct topics is the number of detection of topic transitions, and the top Precision rate is the number of correct topics / N, and the recall rate is the number of topic transitions / N. The proposed method only detected those topic transitions to which gesture labels were given. Therefore, the following two calculation methods were used. Calculation method 1 (equation 2) and calculation method 2 (equation 3), which focused on “topic transitions to which gesture labels were given within the top N” and “the top N of topic transitions to which gesture labels were given.”

No. of correct topics = No. of topic transitions to which gesture labels were given within top N
Precision rate = No. of correct topics / No. of gesture labels given within top N
Recall rate = No. of correct topics / No. of gesture labels given to topic transitions (2)

No. of correct topics = No. of topic transitions in top N within those to which gesture labels were given / N
Precision rate = No. of correct topics / No. of topic transitions to which gesture labels were given

Recall rate = No. of correct topics / No. of gesture labels given to topic transitions (3)

When compared to the exiting method, the proposed method achieved a good Precision rate, while the recall rate decreased. This result was achieved, because those statements that were not accompanied with gestures could not be detected. In addition, in calculation method 2, we have confirmed that both of the Precision rate and the recall rate can achieve good results when only gestures are focused on.

4.2 Comparison Experiment 2

The last half of the debate show was used in comparison experiment 2. Here, those issues of economic stimulus measures and consumption tax were discussed. Table 6 shows the parameters set. The analysis range covered 308 statements, from ID750 to ID1057. The discussion contained the following five topics: the economy, nation's financial condition, increase in the consumption tax, about the Democratic Party, and postal service privatization. 50 gesture labels were given to statements, while topics transitioned 20 times. Figure 4 shows the clustering results.

Table 7 shows the results of the detection of topic transition for comparison experiment 2. The same results with those of comparison experiment 1 were obtained.

Table 4: Parameters of comparison experiment 1

Parameters	Value
Starting IDs	44
Ending IDs	392
No. of topics	6
Core words	Inside the pref., outside the pref.
	Guam, Tinian
	local, mayor
	deterrent force
	Hatoyama, Prime Minister, party heads, chancellor
	the Prime Minister's Office, leak
Gesture labels	150
Topic transition times	39

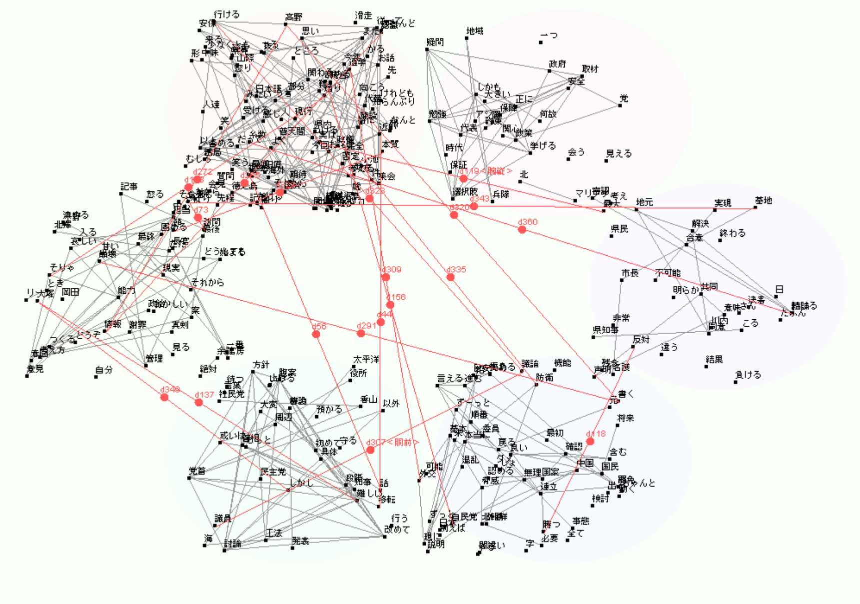


Figure 3: Clustering graph of comparison experiment 1

Upper left: Domestic relocation of the U.S. air base

Upper right: Overseas relocation of the U.S. air base

Left: About the Prime Minister's Office, Right: Willingness of the local residents

Lower left: Prime Minister Hatoyama, Lower right: The deterrent force

Table 5: The results of the detection of topic transitions for comparison experiment 1

		Existing method	Proposing method	
			Calculation method 1	Calculation method 2
Top 20	the number of correct answers	3	1	6
	Precision	15.0%	50.0%	30.0%
	Recall	7.7%	2.6%	50.0%
Top 40	the number of correct answers	9	2	12
	Precision	22.5%	33.3%	30.0%
	Recall	23.1%	5.1%	100.0%
Top 60	the number of correct answers	14	5	
	Precision	23.3%	38.5%	
	Recall	35.9%	12.8%	
Top 80	the number of correct answers	19	7	
	Precision	23.8%	31.8%	
	Recall	48.7%	18.0%	

Table 6: Parameters of comparison experiment 2

Parameters	Value
Starting IDs	750
Ending IDs	1057
No. of topics	5
Core words	Economy
	financial condition
	consumption, tax increase
	the Democratic Party, Manifesto
	postal services, post, postal savings
Gesture labels	50
Topic transition times	20

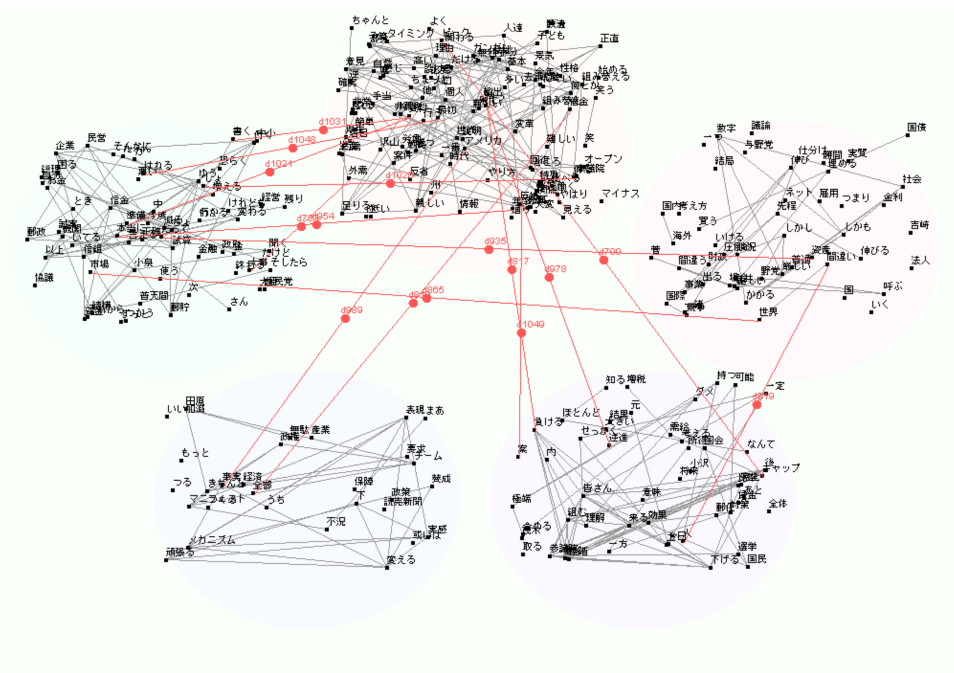


Figure 4: Clustering graph for comparison experiment 2

Upper left: Postal service privatization, Upper center: The economy

Upper right: Nation’s financial condition, Lower left: The Democratic Party

Lower right: Tax increase

Table 7: The results of the detection of topic transitions for comparison experiment 2

		Existing method	Proposing method	
			Calculation method 1	Calculation method 2
Top 20	the number of correct answers	2	0	5
	Precision	10.0%	0.0%	25.0%
	Recall	10.0%	0.0%	62.5%
Top 40	the number of correct answers	7	2	8
	Precision	17.5%	28.6%	20.0%
	Recall	35.0%	10.0%	100.0%
Top 60	the number of correct answers	10	2	
	Precision	16.7%	18.2%	
	Recall	50.0%	10.0%	
Top 80	the number of correct answers	15	3	
	Precision	18.8%	21.4%	
	Recall	75.0%	15.0%	

5. Conclusion

In this research, we proposed a new discussion analysis method that considers not only verbal information, but also non-verbal information. We also performed comparison experiments with the existing method in order to examine the accuracy of detection of topic transition. The proposed method has potential to be used if limited to statements made with gestures.

In addition, scores were given to the ranking function by considering each of the gestures as the same. However, to those who originally use a particular gestures a lot, statements made accompanied by such gestures could have no meaning. In the future, more detailed analysis will have to be conducted. For example, there is a need to discern those who originally make statements with a lot of gestures and those who do not.

References

[Ohsawa 06] Ohsawa. Y., Data Analysis of Chance Discovery (In Japanese), Tokyo Denki University, (2006).

[Maeno 07] Maeno, Y., An Invisible Fixer in an Organization Inferred from Communication, The Japanese Society for Artificial Intelligence, Vol.22, No.4, pp.389-396, (2007).

[Ohsawa 05] Y.Ohsawa: Data crystallization: chance discovery extended for dealing with unobservable events, New Mathematics and Natural Computation, Vol.1, pp.373-392, (2005).

[Nitta 12] Nitta, K., Okada, S., Sugimoto, M., Discussion Analysis Considering Logical and

Emotions, 64th SIG-SLUD, The Japanese Society for Artificial Intelligence, (2012).
[McNeill 01] McNeill, D., Quek, F., McCullough, K., Duncan, S., Furuyama, N., Bryll, R., Ma, X. and Ansari, R.: Catchments, prosody and discourse, Gesture, Vol.1, No.1, pp.9-33, (2001).4

Discussion Analysis Considering Logical Structure and Emotions

Katsumi Nitta Shogo Okada Masaki Sugimoto

Department of Computational Intelligence and Systems Science,

Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

1. Introduction

When a discussion extends for a long period of time, it is a burdensome task to understand the transitions of topics by reading its statement record and to extract those statements that triggered topic transitions. It is very difficult to learn the logical relationships between discussion points and understand the discussion content correctly, especially when the discussion contains multiple points being discussed. Additionally, there exist some factors that make it difficult to understand the discussion according to the statements made during the discussion, such as obscure claims including hidden real intentions.

In the past, a wide variety of methods have been proposed to extract important statements and detect topic transition points by conducting text analysis using a computer in order to make it easier to understand discussion records.

However, in the case of this research we target discussion records of full-scale technical arguments that contain specialized knowledge, such as debates, arbitrations, settlements, and negotiations. It is difficult to understand the transitions of discussion points adequately without having knowledge of various points concerning the main discussion theme and necessary background information regarding the relationships between the points brought out. Furthermore, text analysis is actually insufficient in order to obtain the correct understanding of certain statements that contain hidden intensions.

Given these facts and problems, we conducted research for the purpose of developing a discussion analysis method that takes into consideration background information and emotions. Here, background information indicates discussion points that are presumed in advance regarding the main theme of the discussion, along with the relationships between the discussion points. With regard to emotions, non-verbal information in addition to discussion text information was utilized, such as gestures and voice volume, loud or soft. Non-verbal information is likely to contain information for certain portions of statements that were especially emphasized, when the speaker was emotional or upset.

Utilizing the time-series word clustering method for analysis of non-verbal information, we have proposed in this research a method that observes the relationships between issues of statements and emotional matters. To be specific, we introduced two discussion analysis methods, logic analysis and co-occurrence analysis of words. Logic analysis provides preliminary knowledge regarding the main

theme of the discussion, and it offers opinions and objections that can be presumed in advance, along with logical relationships on the grounds of such opinions. Co-occurrence analysis of words observes that in the record of a discussion that was actually conducted, what kind of words appear simultaneously in statements, in order to extract those statements that make topics transition or triggered topic transitions. When co-occurrence analysis is conducted, it is assumed that the logic analysis results are referred to for determining the grading size of co-occurrence level analysis.

Chapter 2 describes how to organize and summarize the points of the theme. Chapter 3 gives an explanation regarding transitions in points to be discussed focusing on co-occurrence of words in the statement records and an analysis method for important statements. Finally, chapter 4 describes how to apply this method to non-verbal information analysis, and application examples.

2. Logic Analysis on the Main Theme

Logic analysis on the main theme is supposed to express the relationship among points to be discussed regarding the main theme in the Toulmin diagram format in advance. For example, in the issue of the relocation of the U.S. air base in Okinawa, major points to be discussed can be listed in advance, such as the burden for the residents of Okinawa, the need of the U.S. military base, the need of locating the U.S. base in Okinawa, and the need of relocating the U.S. base to another prefecture out of Okinawa. When it comes to the issue of reactivating nuclear power plants that are currently in a dormant state, major points to be discussed can be predicted in advance. Such points should include the possibility of the occurrence of huge earthquakes, a stable electric power supply, and the possibility of alternate measures for electric power generation.

As an example, Figure 1, the diagram format, shows part of the major points to be discussed, which could be predicted in advance under the main theme of the issue of reactivating nuclear power plants. In this figure, the points to be discussed are expressed by nodes, while the relationships between the points are expressed by links. One-way links indicate support relationships, while mutual links indicate conflict relationships. We can consider that when the points to be discussed are preliminarily listed, discussion transitions on these points to be discussed, and individual statements correspond to the refinement of these points. Even though the list contains the same words, the points to be discussed transition in various ways depending on the way the discussion progresses, for instance, particular points are discussed in depth, the sides of the discussion are changed, or the points to be discussed transition inversely.

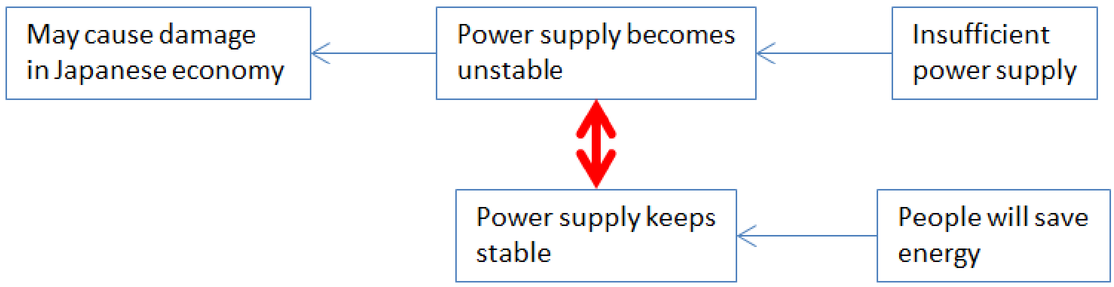


Figure 1: Organization of points to be discussed

3. Analysis on Co-Occurrence of Words

3.1 KeyGraph and Word Clustering Analysis

The group of Ohsawa and Maeno has proposed a method of analysis on points to be discussed by utilizing KeyGraph and data crystallization. In addition, they have proposed a word clustering method and another analysis method by means of data crystallization [2], as complementary methods for analysis based on KeyGraph. By doing so, they showed that these methods can be utilized for extracting the speaker's hidden intensions. Based on this word clustering method, we tried to enhance this method in order to improve the extraction accuracy of hidden intensions and to extract important statements. Where the discussion record is considered to be a set of s1, s2…, and each statement Si is considered to be a set of words that appeared, {w1, w2…,}, the method proposed by Maeno et al. defines the distance d(wi,wj) between each word as the reciprocal of the Jaccard index. Next, all words that appeared in statements are clustered into the given number N, by utilizing the K-medoids method (Figure 2).

When each word is expressed with a node and words having a high Jaccard index are connected with links, a graph that consists of n islands (clusters) can be obtained. Each cluster is probably considered to be a single topic.

Next, a dummy word di is inserted into each statement si(i=1,2,…). Where a certain statement Si contains multiple clusters inside, dummy words and the representative words of both clusters are connected and added to the graph. When these dummy words appear on the graph as nodes, they are referred to as dummy nodes. The appearance of these dummy nodes suggests that the statement E that corresponds to these nodes refers to several topics. This indicates that other topics are mentioned during the statement about a certain topic, or a topic is guided to transition to another topic.

The use of dummy nodes has the possibility of discovering the characteristics that are not expressed on the surface of the statement record. For example, Maeno has shown the possibility of extracting the hidden intentions contained in a statement by utilizing dummy nodes. This is because topics that attract attention and interest can be predicted by making statements that contain related words even without making clear statements.

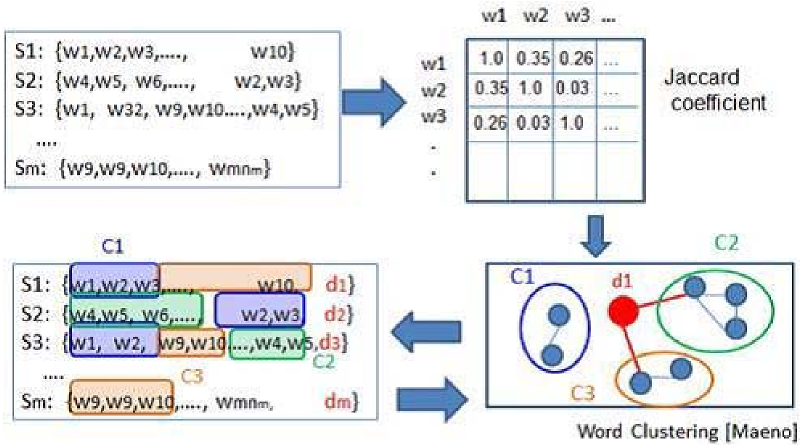


Figure 2: Word clustering and dummy nodes

3.2 Word Clustering Considering Time Series

Word clustering is an effective method for analysis of topic transitions based on the discussion record. However, this method has the issue that the clustering precision might decrease when the discussion extends for a long period of time and contains a lot of topics, along with complicated topic transitions. For example, depending on words, it could be natural to classify words into different clusters between the first half and the last half of the discussion. However, the above-mentioned clustering method can only classify words into the same cluster through the entire discussion.

Given this issue, we proposed two ways to enable this word clustering method to handle the passage of time. One method is, when the Jacard index is calculated, to consider not only co-occurrence within the same statement, but also co-occurrence between adjacent statements. The other method is to divide the discussion record at each point where topics make a significant shift in order to re-cluster words according to each individually divided section.

In this research, we proposed the latter method. By applying the word clustering method for the entire discussion record, the words that appear are divided into a given number of clusters (Figure 3). Next, a histogram, which shows how words appeared in each cluster as time passed, is obtained.

This histogram shown as bar charts indicates each of clusters. When there is a point where two lines clearly cross, this point is determined to be where topics made a significant shift. Before and after this point, the discussion record is divided into two sections, and then the word clustering method is applied to each of these divided sections.

Afterwards, repeating this process divides the discussion record in a hierarchical way.

Just as described above, those topics that are discussed partly within each section can be made obvious by analyzing each individual section. In addition, dummy nodes that make strides over both

sections can be extracted, by overlapping a few statements existing around the section border of both sections when sections are divided. Compared to those dummy nodes existing within the sections, those dummy nodes making strides over multiple sections can be considered to be statements that are related to the shift of topics.

For example, Figure 4 shows part of the analysis results of text data extracted from the debate of party leaders that took place between Prime Minister Aso and the Democratic Party Leader Hatoyama in 2009 (the word clustering results of the first two sections separated from the entire discussion record). In this debate, those points such as economic recovery measures, child rearing policies, the Okinawa base issue, the nation's financial condition, bureaucrat-based politics, and the pension issue were discussed. Of these points, these four points consisting of economic recovery measures, child rearing policies, the nation's financial condition, and the pension issue have a deep connection. Therefore, points that are related to these above-mentioned points were often mentioned

In the first section, economic recovery measures were mainly discussed, and bureaucrat-based politics was discussed in the second section. In the third section, the pension issue was discussed. Dummy node d13 that links the first section with the second section exists in this figure, which corresponds to the following statement.

"Neglecting wasteful spending to the critical extent and getting into debt, and then ending up raising the consumption tax, these things have been done by the current ruling party. I think anybody can conduct such politics. Once again, I would like to say that we must stop such politics in which the people need to carry these burdens. I think this situation has been caused by such a bureaucrat-based government. When it comes to this matter, I'd like to say one more thing. Mr. Aso said that they would totally ban the practice of reemployment of bureaucrats in his opening speech. However, I've heard that LDP removed the portion which describes prohibiting the reemployment of bureaucrats from their manifesto on the day they presented it. I can't understand why they did that. I'd like to hear from you as to why the LDP deleted that portion from the manifesto."

The point related to reemployment of bureaucrats was discussed after this statement was made by Leader Hatoyama. Therefore, this statement led the topic to transition from economic recovery to bureaucrat-based politics.

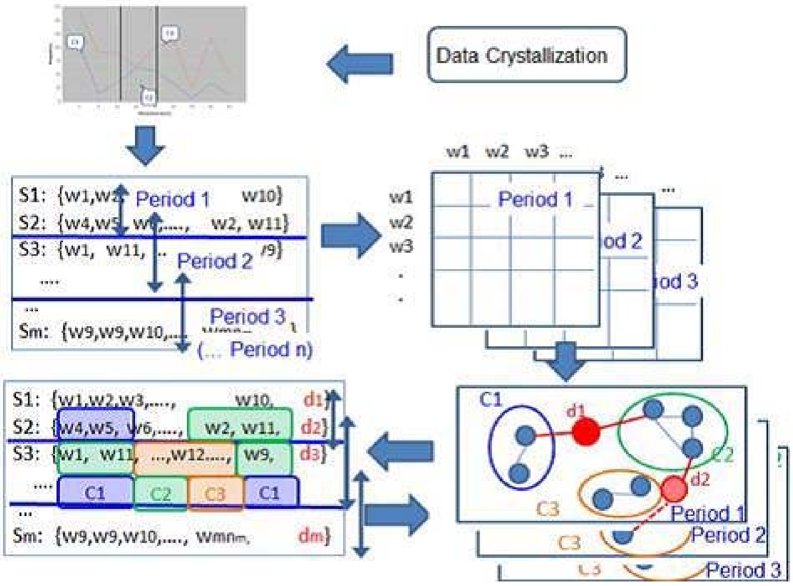


Figure 3: Time-series word clustering

4 Word Clustering Introducing Non-Verbal Information

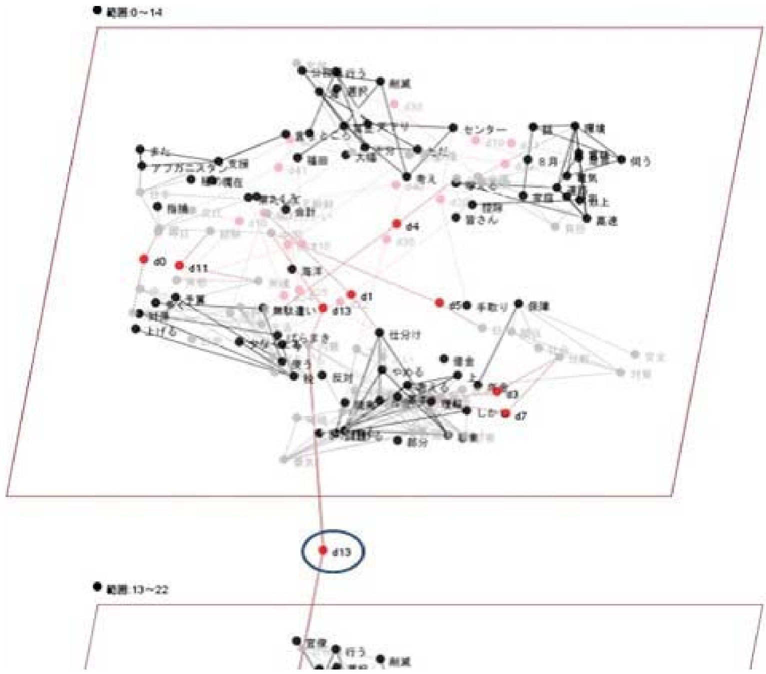
The previous section showed the possibility of extracting important statements, such as those statements that trigger a shift of topics, by focusing on dummy nodes. However, dummy nodes are determined by whether words from multiple clusters are included in one statement or not. Therefore, dummy nodes are easily affected by noise, and it cannot be discerned whether multiple topics with what kind of relations are included.

Recently, however, the video recording of court trials has been considered, mock trials conducted in the departments of law at universities have been recorded, and debate programs have been broadcast on TV. Amid such circumstances, not only text information, but also non-verbal information has been extracted from discussion records. In this research, we attempted to support the interpretation of dummy nodes, by applying such non-verbal information to the word clustering method.

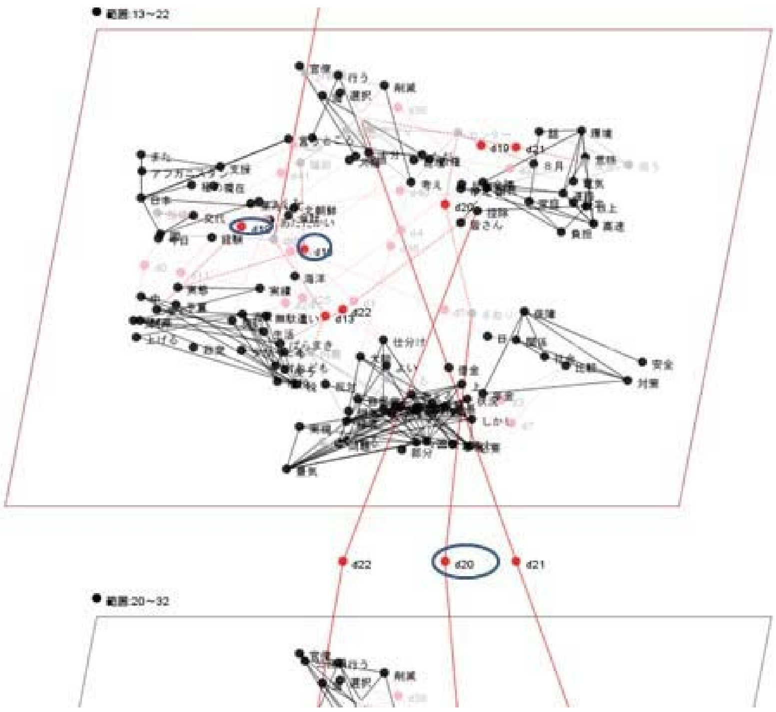
4.1 Overview of Word Clustering Introducing Non-Verbal Information

(1) Extraction of non-verbal information and labeling

We targeted discussion data; therefore, speakers made most of their statements while sitting. We focused on three parts of the upper body, the head, the trunk, and the arms, in order to observe whether speakers would show characteristic movements while making statements during the discussion. Afterward, labeling which is shown in Table 1 was done. This information was tagged to statement text and expressed in an XML format.



(a) Period 1



(b) Period 2

Figure 4: Analysis of the debate of party leaders

Table 1: Labeling of non-verbal information

Head: Downward, forward, and nodding

Trunk: Rightward, backward, leftward, forward, and back and forth

Hands and arms: Putting hands on the chin, moving hands sideway, moving hands up and down, crossing arms, and lacing fingers

(2) Analysis of topic transitions by word clustering

When the word clustering method considering time series is utilized, as shown as 1/2 and 3/4, those labels that express gesture information can be considered to include 3/4 kinds of labels. The first method is to make words co-exist with labels, as shown below.

$$Si=\{ wil,\ldots,wim, ak,di \}$$

In this method, words and gesture labels become the target of clustering equally. Therefore, conducting time-series word clustering provides information as to which point is being discussed when a certain gesture is expressed.

The other method is to consider gesture labels as the attributes of words.

$$Si=\{ wil,\ldots,wim, d(aj,ak) \}$$

Using this method produces the same clustering result with the method without preparing gesture labels. Gesture labels are utilized when the meaning of dummy nodes is interpreted and evaluated. For example, when the gesture label “nodding” is given to dummy nodes, we interpret these dummy nodes by considering that this statement was possibly made by agreeing with the opinion of the opposing speaker.

The next section describes an analysis example by using the second method.

4.2 Analysis Example

This section shows an example of analysis on the record of statements from a debate TV show, “Asamade Nama TV” (A Live Telecast until the Morning), on air in 2011. This TV show had 14 participants including the program presenters, discussing social issues of the relocation of the U.S. air base in Okinawa, economic stimulus measures, and the consumption tax.

With respect to the relocation of the U.S. air base, the possible points to be discussed and part of the relationships with the points can be expressed as shown in Figure 5.

Table 2 shows gestures (including whether the participants raised their voice) that were observed in the show.

Each participant has their own habits in their gestures. To those who have their own unique gestures and use them a great deal, statements made accompanied by such gestures could have no meaning. In contrast, the statements made by those with seldom used gestures do have meaning.

Figure 6 shows the clustering results of words in the middle part of this TV show. This figure

shows such points discussed as politics in general, the deterrent force, the words and deeds of the Prime Minister, economic stimulus measures, and the policies of the Democratic Party. In this figure, since the attribute “in front of the trunk” (forward) is given to dummy nodes 304 and 310, the area around these statements are observed. As shown below, the periphery of these dummy nodes are comparatively emotional points. Politics in general, the attitude of ministers, was discussed immediately before this dummy node, whereas the points of politics in general and the policies of Democratic Party were mentioned in statement 304. In statement 310, the points about politics in general and defense including the U.S. base issue were mentioned.

ID297: (Omitted) Why do ministers pretend ignorance? (Putting the hand on the chin)
ID298: Well, I personally don’t know if they really pretended ignorance. (Crossing the arms)
ID299: They pretend not to notice.
ID300: I feel pity for them if they are regarded that way, while they’re not here. But actually, everybody knows that the problem is deep-rooted. Amid the situation where every ministry has already been determined, it will become confusing if other people try to take the lead in connection with the problem I think… (Omitted) (Crossing the arms)
ID301: Do not mess around.
ID302: I feel if you have ever wondered about messing around. (Bringing hands together)
ID303: They pretended, right?
ID304: (Omitted) Doesn’t Democratic Party have little interest in the foreign and defense policies from the beginning? (In front of the trunk)
ID305: Oh, that’s not the case.
ID306: No, no, because the members do not intensely discuss such issues. (In front of the trunk)
ID307: No, I don’t agree with that. That’s not the case at all. (Arms on the sides)
ID308: OK, then when we have elections, I don’t think you will include any words or expression about foreign and defense in your Manifesto. Your Manifesto does not have such words or expressions, right?
ID309: That’s not true. (Omitted)
ID310: Oh yes, it’s true. Only a small part, just a small part on “diplomacy” is included. (Omitted) Members gathered together just for the purpose of winning the election campaign. But problems have actually come to the surface, because you haven’t conducted any intense discussion about the basic national policies including the core issues, such as foreign affairs and defense. (In front of the trunk, with a loud voice)
ID311: (Omitted) We were doing what we should do in order to take charge of the administration during the period when we were the opposition party. And we’ve been taking control of the government for eight months, and also accepting stern criticism. As the ruling party that is taking

charge of the administration, it is nonsense to say that we are not interested in foreign diplomacy. (Bringing hands together)

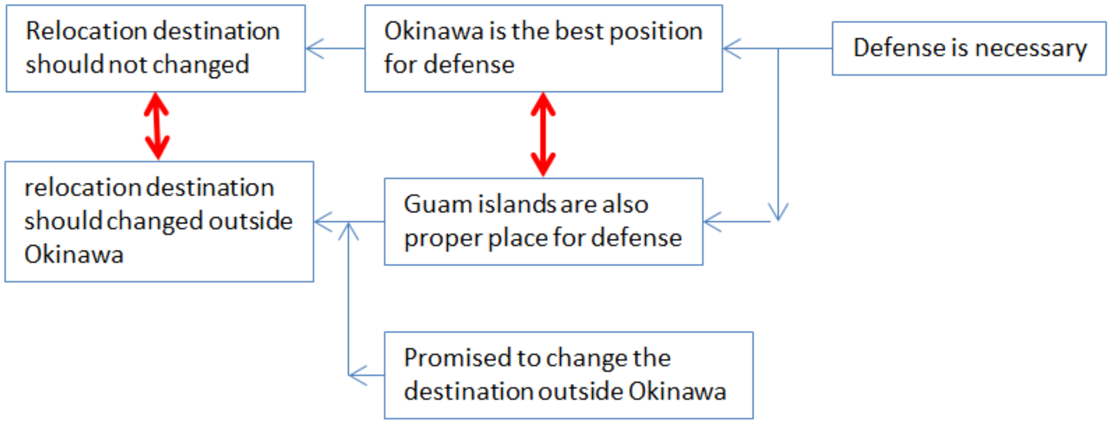


Figure 5: Example of the points to be discussed regarding the relocation of the U.S. air base

The relocation destination must not be changed; Okinawa is the best place for exercising the deterrent force; The deterrent force is needed; The relocation destination should be changed to the outside of Okinawa; Locating the base in Guam is still sufficient for demonstrating the deterrent force; Promised to relocate the base out of Okinawa

Table 2: Gestures of the participants in the discussion

Speaker	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Downward	0	0	0	0	0	0	2	1	0	0	0	0	4	1
Forward	3	1	0	2	9	2	1	1	0	5	0	0	1	0
Nodding	0	0	0	0	1	1	6	2	0	0	0	1	0	1
Rightward	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Backward	0	0	0	0	0	0	0	0	0	2	0	0	1	4
Leftward	0	0	0	0	0	0	0	0	0	1	0	1	0	0
Forward	9	1	4	1	1	1	2	4	0	24	1	0	4	3
Back and forth	1	0	0	0	0	1	0	0	0	0	1	0	0	0
Hands horizontal	0	0	0	2	0	0	3	0	0	0	0	2	1	0
Hands vertical	0	0	0	2	0	3	13	10	3	19	0	3	4	5
Folding	3	2	0	7	0	0	0	0	0	0	0	0	1	3
Bringing together	9	0	0	18	0	0	2	0	0	0	2	2	0	0
Loud	1	0	0	0	0	0	1	0	0	17	0	0	0	0
Total	26	4	4	33	11	8	30	18	3	68	4	9	16	17

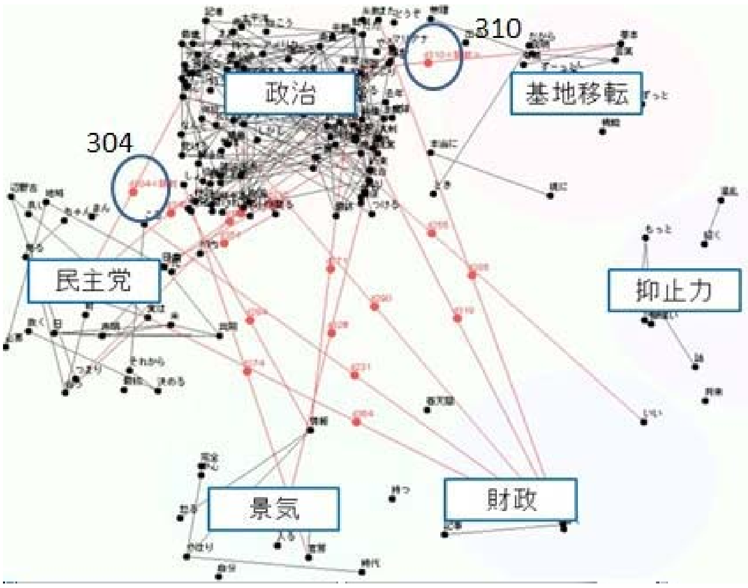


Figure 6: Word clustering graph of the debate program

5 Conclusion

In this report, we introduced the extraction of the points to be discussed from the main theme, and a method of analysis on co-occurrence based on the record of statements. Afterward, the possibility of applying this method to analysis of non-verbal information was described. The effects of this method are now being studied and confirmed by applying it to various experiments performed. However, we consider this method to be useful and effective for extracting important statements.

In the future, as a dummy node attribute, we consider adding not only non-verbal information, but also the grammatical characteristics of the statement as labels.

Acknowledgement

We would like to express our appreciation here for Professor Yukio Ohsawa (Tokyo Univ.) and Ph.D. Yoshiharu Maeno (AI Group) for their fine instruction regarding word clustering techniques.

References

[1] Ohsawa, Y., eds, Chance Discovery in Real World Decision Making, Springer-Verlag, 2006.

[2] Maeno,Y., Invisible Structure of Decision Making Discovered from Communication, Ph.D, Thesis, 2007.

[3] Maeno,Y., Ohsawa,Y., Human-computer interactive annealing for discovering invisible dark events, IEEE transactions on Industrial Electronics 54, pp. 1184-1192,(2007)

[4] Maeno, Y. Nitta. K, Ohsawa, Y., Reflection of the agreement quality in mediation, Proc. 3rd

International Workshop on Juris-Informatics, pp.73-82, (2009).
[5] Nitta,K., et al., Extraction Support System Using Word Clustering and Data Crystallization, Proc. 3rd International Workshop on Juris-Informatics, pp.95-106,(2009).

Multimodal Discussion Analysis based on Temporal Sequence

Katsumi Nitta

Interdisciplinary Graduate School of Technology and Science,
Tokyo Institute of Technology

Abstract. This research proposes a novel method for analysis of discussion record. One of the important features of our approach is to use both a logical analysis method and a word occurrence analysis method. A subject of discussion is analyzed and important issue factors are listed before the discussion starts. The logical analysis method describes the structure of the discussion referring to the issue factors. The word occurrence analysis method recognizes key topics and key utterance by observing utterances and nonverbal information such as action, facial expressions and so on.

1 Introduction

Discussion plays an important role in the resolution of disputes such as negotiation, moderation and arbitration. Discussion is modeled as exchanges of *arguments* on a specific topic. An argument is a pair of conclusions and their grounds which supports each conclusion. During discussion, an argument may be attacked by a counterargument. By exchanging arguments and counter arguments, discussion becomes more detailed and more complex. When a discussion includes a lot of topics (issue points) and they are related to each other, it sometimes becomes hard for participants to capture the whole structure of the discussion to understand which issue points are used to reach a consensus, and which arguments defeated other counter arguments. In such cases, a discussion support system which visualizes the structure of arguments and shows various features to evaluate the discussion skills will be helpful.

To support the analysis of discussion, a lot of research has been conducted so far. For example, some research done has represented the logical relationship among arguments in the form of a diagram, and has analyzed structure of a discussion [Reed 04]. Another research represents features of a discussion as a set of propositions, and estimates its conclusion by searching for similar cases from the past [Aleven 97]. Other research has focused on the statistics of the utterances that occurred during the discussion, and tries to find time points where topic change occurred [Ohsawa 06].

The researches done in the past have showed a useful effect for analyzing some aspects of discussion. However, they have been insufficient in anlyzing not only the structure of discussion, but also discussion moderation skills. For example,

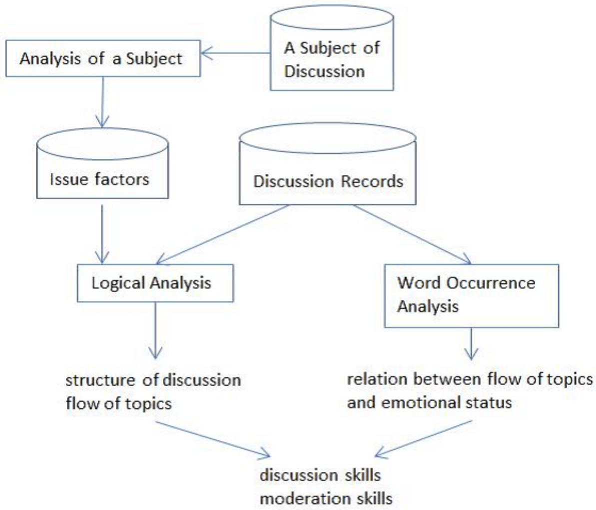


Fig. 1. Discussion Analysis Using Two Methods

when a discussion subject is decided, several topics to be discussed are estimated before the discussion starts. The subject includes several topics, and each topic is composed of several issue points. If the discussion is well moderated, these topics are discussed effectively. However, if the discussion skill is not of a high level, some important topics may be skipped or some topics may be discussed repeatedly. The emotional status of participants is also important for analyzing discussion moderation skills. If some participants cross their arms while speaking, he may be showing that he is irritated. In such a case, the chair person should change the topics or should take a coffee break considering the atmosphere of the discussion. Discussion moderation skills can be analyzed by extracting the logical structure and the emotional status from the discussion records and by comparing the discussion records with other discussion records whose subjects are the same. However, traditional analysis tools are not adequate to analyze discussion skills.

The objective of this research is to propose a novel method which supports the analysis of discussion record. One of the important features of our approach is to use both a logical analysis method and a word occurrence analysis method [**?**] [Maeno 06] [Nitta 09]. Fig. 1 shows the relation between the two methods. A subject of discussion is analyzed and important issue factors are listed before the discussion starts. The logical analysis method describes the structure of the discussion referring to the issue factors. The word occurrence analysis method recognizes key topics and key utterance by observing utterances and nonverbal information such as action, facial expressions and so on.

In Section Two, the logical analysis method is introduced. In Section Three, the temporal word clustering method and its extension to nonverbal information are introduced.

2 Logical Analysis of Discussion Records

A discussion is held based on the exchange of messages made by natural language. Although the same content is discussed, wording of the utterances and expressions vary depending on each individual speaker. For this reason, to compare several discussion records, we need to decide common factors to describe arguments which appear in the discussion.

We consider here that an argument is described using a proposition that indicates a fact or claim, and we call this proposition an *issue factor* (in this paper, we call it a *factor*). There exist a relationship where the establishment of one's factor supports (serves as the basis for) the establishment of the other one's factor, and a relationship where the establishment of one's factor attacks (conflicts with) the establishment of the other one's factor.

The following shows an example of the factors. In this case, f1 is the base of f3 (f3 holds because f1 holds), f2 is the base of f4 (f4 holds because f2 holds), while f3 conflicts with f4.

f1: The product sold was out of order.
f2: No malfunction was found in shipping.
f3: The seller is at fault.
f4: The seller is not at fault.

These relationships among factors can be expressed in an *issue graph* that is shown in Fig. 2. In this graph, each node corresponds to a factor, and a solid arrow shows a support relation between two factors and a dotted arrow shows an attack relation between two factors. This graph can be drawn when the main subject of a discussion has been determined before the discussion starts.

When a discussion starts, issues move onto this issue graph. Even though in discussions with the same subject, the transition of issues significantly varies depending on each individual speaker. By observing this transition, discussion moderation skills are evaluated.

To extract issue factors from utterances, morphological analysis should be conducted on the utterance messages and issues are estimated by utilizing combinations of words that occurred. We have proposed a machine learning method that discerns groups of words used for extracting issue factors from utterance records according to issues. Where multiple issue factors are extracted, an argument (a pair of conclusions and reasons) could be built depending on the combination of multiple issue factors. For example, when two issue factors, f1 and f3, are extracted and where there exists a relationship in which f1 supports f3 on the issue graph, this situation can be considered to represent the argumentation that "it is f3 because of f1."

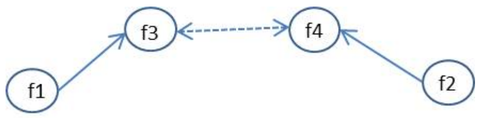


Fig. 2. Issue Graph

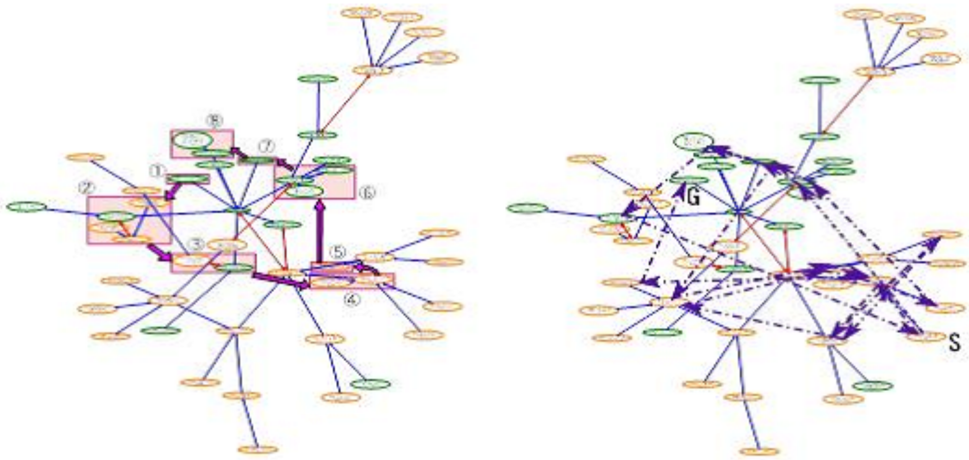


Fig. 3. Issue Graph and Topic Flow

As we described, even if the structure is the same, the process of discussion varies. For example, Fig.3 shows two flows of the topic of the discussion of the citizen judge system. The left figure is a topic flow by the lawyers, and the right figure is that of students. Lawyers topic flow is very effective because the same factor is not raised twice. On the contrary, right figure shows the same topic is brought up several times, which means this discussion is ineffective [Sato 11].

3 Word Clustering Analysis

3.1 Basic Word Clustering Method

Maeno and Ohsawa regarded a discussion records as a set of utterance S1,S2,··· Sm, and each utterance as a set of words {w1,w2,···,wn} which occurred in the utterance [?] [Maeno 06]. Distance between any two words (wi and wj) is defined using the Jaccard coefficient.

Given the intended number of clusters, all words are clustered using K-medoids method (Fig. 4). In Fig. 4, all words are clustered into three clusters (C1, C2 and C3). Each cluster is represented as a set of nodes and links. Each

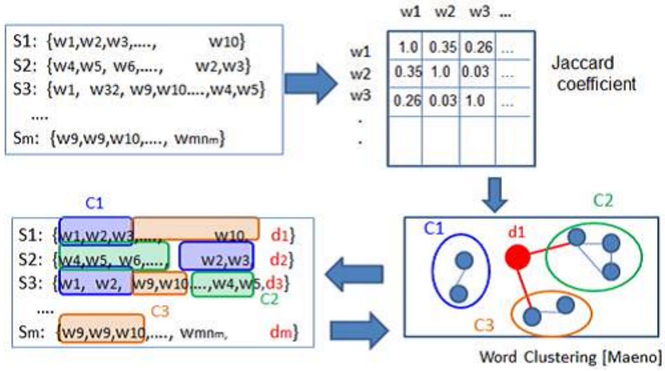


Fig. 4. Word Clustering Method and Dummy Nodes

node corresponds to a word, and each link shows that Jaccard coefficient between these two words is a high score. We think that a cluster corresponds to a topic, so during discussion the focal cluster moves according to the change of topics.

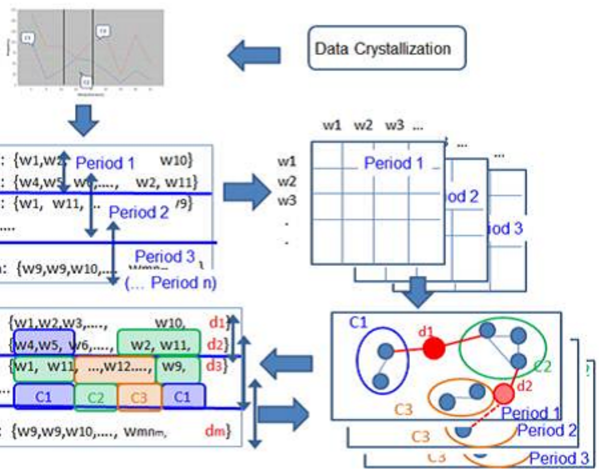


Fig. 5. Temporal Word Clustering Method

After the clustering phase, in each utterance, a dummy node is inserted. A dummy node is different from other nodes because it doesn't represent a word, but it corresponds to an utterance. For each utterance, a ranking function is calculated. A ranking function measures the numbers of clusters which occurred

in the utterance. If the value of a ranking function of an utterance is more than the threshhold, and if a cluster Ci is the largest cluster and Cj is the second largest cluster, then from a dummy node to the representative nodes of the Ci and Cj, we make links. In the example in Fig. 4, a dummy node d1 combines C2 and C3 which means that C2 and C3 are large clusters in the utterance S1.

By interpreting the role of dummy nodes, we can extract various kinds of information. For example, some dummy nodes correspond to utterances where one topic is spoken by reffering to another topic. Another dummy node corresponds to utterances where one topic is spoken, wanting to change the topic to another one.

3.2 Temporal Word Clustering Method

When the discussion record is small in size, the original word clustering method works well. However, when the size becomes bigger, precision of word clustering decreases because the role of each word may change during the discussion and because a small cluster may be absorbed into a bigger one. To cope with the problems, we devised the temporal word clustering method. This method divides the discussion records into several parts at the points where the topic changed a lot as follows.

At first, for the total discussion record, the original word clustering method is applied. Then, we count the number of words for each cluster in chronological order, and make a histogram in which each cluster is represented by a line. In the histogram, we find time points where two lines cross, and at these time points, the record is divided into several sub records (discussion periods). Then, for each sub record, the above process is applied hierarchically. As a consequence, if a discussion record is divided into N sub records, the output of the temporal word clustering becomes N clustering graphs (Fig.5). As we divide the discussion record leaving some overlaps, dummy nodes in these overlaps may combine two clusters which belong to adjacent word clustering graphs. These dummy nodes correspond to utterances which caused the change of topics.

3.3 Multimodal temporal word clustering

Dummy nodes correspond to utterances which include more than one topic. By observing these topics, we can estimate important changes in topics. However, when there are utterances that include more than one topic, their interpretation is not easy.

Sometimes, discussion records are given in the form of a movie file, and we can observe facial expressions, actions such as crossing arms, nodding, hand waving, falling forward and so forth. Such nonverbal information is useful to estimate the emotional state of individuals while they are speaking. Therefore, by combining a temporal word clustering method and multi modal information, we can extract more detail information which supports the interpretation of dummy nodes.

Nonverbal information in the discussion records can be analyzed as follows. At first, we observe the movie file and extract nonverbal information listed in

Table 1. The extracted information is tagged according to each utterance by using iCorpus Studio [4] , and is saved in an XML format as shown below.

```
<Utterance id="15" Speaker="Yamada" Issue="F3"
   Head="Tilt" Body="Straight" Arms="Crossed">
    Contract dissolution will not be granted
</Utterance>
```

Then, each utterance Si is represented as follows.

$$Si = \{ w1,w2,\cdots,wn;\ a1,\ a2,\cdots,\ am;\ di \}$$

Here, "w1, w2,⋯, wn" are words which appeared in Si, "a1,a2,⋯,am" are the name of the speaker and nonverbal information which appeared in Si, and "di" is a dummy node.

Table 1. Type of nonverbal information

Body Part	Type
Head	Straight, tilting to right, tilting to left, upward, downward, nodding, sticking out, leaning
Body	Straight, tilting forward, tilting backward, tilting to right, tilting to left, swinging
Arms	Crossed, putting the hands forward (vertically), putting the hands forward (horizontally), touching the chin, hiding the mouth, touching the head

We show an example of a TV discussion program, " Asa made nama TV. " In this program, 14 people joined and discussed about 6 problems with the Japanese Government such as the manifesto of Japanese Democratic Party, Okinawa's military bases, Defense of Japan, Economical stimulus policy, Government 's Finance, and so on. Participants are composed of economists, journalists, militarists, psychologists and representatives.

During the 4-hour discussion, several actions were observed as seen in Table 2. In this table, from Head1 to Head3 correspond to positions of the head such as 'downward', 'straight' and 'nodding', respectively. From Body1 to Body5 mean positions of the body such as 'tilting right', 'tilting backward', 'tilting left', 'tilting forward' and 'swinging', respectively. From Arms1 to Arms6 means movement of arms such as 'touching chin','putting hand forward', 'putting hand aside', 'putting hand vertically', 'crossing arms' and 'crossing fingers', respectively. In this table, we can observe several actions of speakers. For example, Mr. Yamagiwa and Ms. Kayama showed several actions while they were speaking. On the contrary, Mr. Morimoto and Mr. Uesugi showed little actions. If these actions are affected by certain emotional status, then by observing the relation between these actions and topics, we can estimate the role of each one's utterance to a greater degree.

Table 2. Actions during discussion

Speaker	# speak	Head1	Head2	Head3	Body1	Body2	Body3	Body4	Body5	Arm1	Arm2	Arm3	Arm4	Arm5	Arm6
Tahara	433	0	3	0	0	0	0	9	1	10	0	3	9	1	0
Itokazu	33	0	1	0	0	0	0	1	0	0	0	2	0	0	0
Uesugi	63	0	0	0	0	0	0	4	0	0	0	0	0	0	0
Ohtsuka	99	0	2	0	1	0	0	1	0	0	1	2	7	18	0
Katsuma	33	0	9	0	0	0	0	1	0	1	0	0	0	0	0
Kayama	13	0	2	0	0	0	0	1	0	1	1	3	0	0	0
Kawauchi	106	2	1	6	0	0	0	2	0	0	3	13	0	2	1
Motegi	77	1	1	2	0	0	0	4	0	1	0	10	0	0	0
Morimoto	51	0	0	0	0	0	0	0	0	0	0	3	0	0	0
Yamagiwa	61	0	5	0	0	2	1	24	0	0	0	19	0	0	17
Yoshizaki	20	0	0	0	0	0	0	1	0	0	0	0	0	2	0
Takahashi	58	0	0	1	0	0	1	1	0	0	2	3	0	2	0
Takano	58	4	1	0	0	1	0	4	0	0	1	4	1	0	0
Koike	78	1	0	1	0	4	0	3	0	8	0	5	3	0	0

There are two ways to analyze the nonverbal information. Method 1 is to consider a multimodal label as a word, and apply the temporal word clustering method to the following basket.

si = { w1, w2, ⋯, wn, a1, a2, ⋯, am }

In this method, each nonverbal label belongs to one of the clusters, which means that such a label is related to the cluster the most. For example, if the nonverbal label is the speaker 's name, it means such a topic that he talks about most. If the nonverbal label is an action such as " crossing hands ", it means that " crossing hands " are observed during the discussing such a topic.

Method 2 is that we treat multimodal labels as attributes of the dummy node as follows.

di(a1, a2, ⋯, am)

In this method, the result of the temporal word clustering with nonverbal information as the same as that of temporal word clustering without nonverbal information. However, to each dummy node, several other forms of information such as the speaker 's name, action labels which is observed while speaking and so on are attached. This information is then used to interpret the meaning of the dummy node.

Fig. 6(a)-(f) shows the results of Method 1. There are 6 figures, because the discussion is divided into 6 periods. In these figures, several kinds of information such as names of speaker, actions, and roles of utterance appear as nodes. As Mr. Tawara is the chair person, his name appears in various clusters throughout these figure, and there are a lot of dummy nodes from Mr. Tawara connected to other clusters. To these dummy nodes, the roles of the utterance are attached. According to this information, most of what Mr.Tawara 's speaks are YES/NO questions, open ended questions, agreements and so forth.

Several action labels appear on the upper left cluster which corresponds to the topic of " violating manifesto. " This topic is a very general one and the other 5 topics are closely related to this topic. Therefore, even if the other topic is the on-point one, the topic " manifesto " is often referred from this topic. In Period 1, " Transferring Okinawa's military bases " is discussed. While 4 participants mainly talked about this topic, most other participants talked " violating manifesto. " Most nonverbal labels occurred in these two topics. In Period 4, " financial

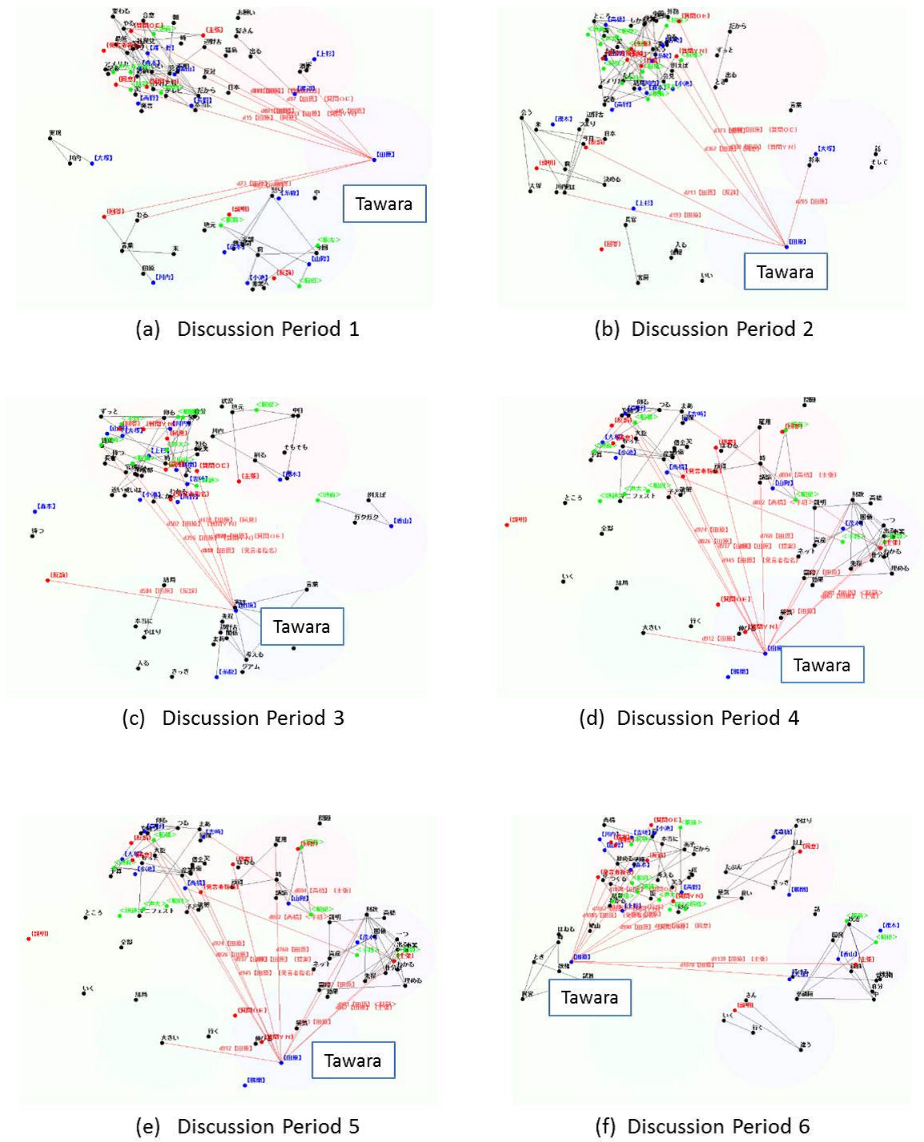


Fig. 6. Word Clustering Graphs

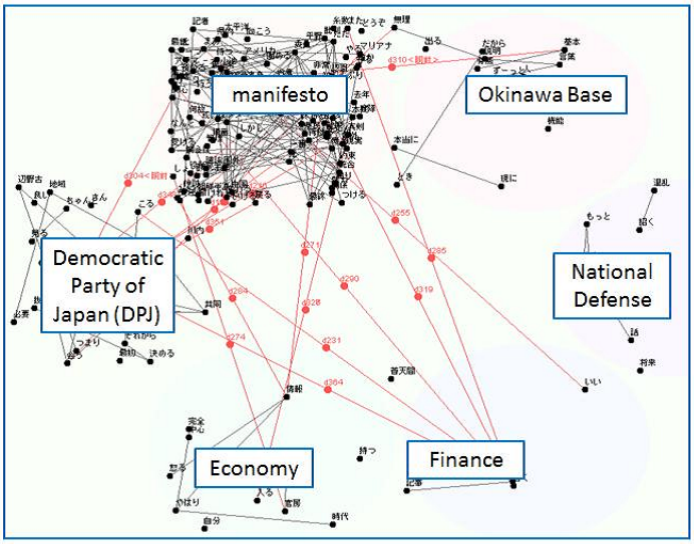


Fig. 7. Result of Multimodal Temporal Word Clustering

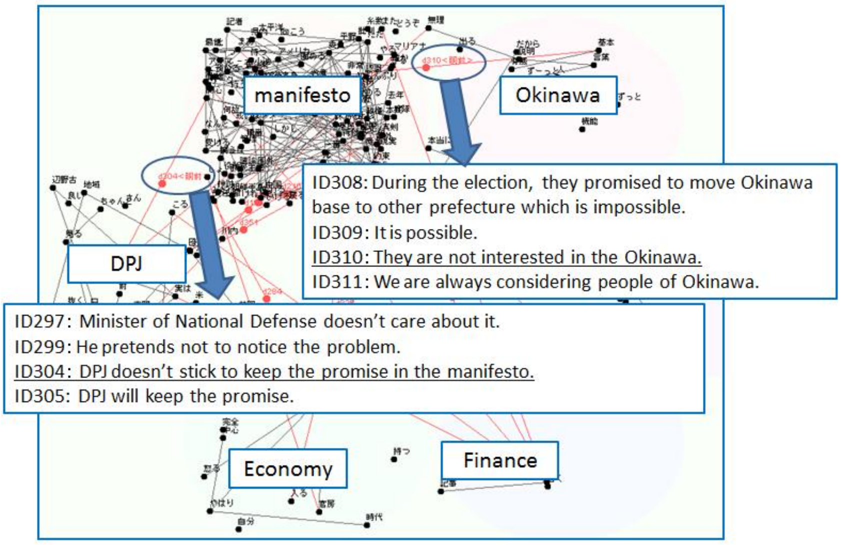


Fig. 8. Dummy Nodes 304 and 310

problem ” is mainly discussed. In this period, most nonverbal labels appear in “ violating manifesto ” topic because this topic tends to be the most exciting one.

Now we will show a more concrete example. Fig.3.3 shows the result of Method2. Six clusters corresponds to six topics such as ”keeping promises described in manifesto”, ”transferring military bases in Okinawa to other prefectures”, ”importance of the relationship with the United States for national defense”, ”financial deficit of the Government and consumption tax”, ”stimulating the economy” and ”fraction of DPJ.” In this figure, several dummy nodes appear. We focus on two dummy nodes 304 and 310 as seen in Fig. **??**. To these dummy nodes, multimodal label ”tilting forward” is attached, which means these utterances were made emphatically. In Fig. 3.3, utterances around these nodes are shown. Actually, these utterances played an important role to bring about change.

4 Conclusion

Here we have shown a novel method for analyzing discussion records. This method uses both logical analysis of the discussion subject and statistical analysis (a temporal word clustering method) of discussion records. By integrating these two methods, we can evaluate discussion moderation skills and the analyzed data that are reused in other discussion where the subject is the same. Furthermore, we showed a method for nonverbal information by extending the temporal word clustering method.

Acknowledgement
This research was supported by AFOSR/AOARD grant no. 114061. I thank AFOSR/AOARD for their support. Also I thank Professor Yukio Ohsawa (The University of Tokyo) for his technical advice about the word clustering method.

References

[Reed 04] Reed, C., Rowe, G.: Araucaria: Software for argument analysis, diagramming and representation. International Journal on Artificial Intelligence Tools 13(4), 961 - 979 (2004)

[Aleven 97] Aleven, V., Ashley, K.D.: Teaching CaseBased Argumentation through a Model and Examples: Empirical Evaluation of an Intelligent Learning Environment. In: Proceedings of AIED 97 World Conference. pp. 87 - 94. IOS Press (1997)

[Ohsawa 06] Ohsawa, Y. eds: Chance Discovery in Real World Decision Making , Springer- Verlag, 2006

[Maeno 06] Maeno, Y. et al. : Crystallization highlighting hidden leaders, Proc. IPMU, (2006).

[Maeno 09] Maeno, Y. Nitta, K., Ohsawa, Y. : Reflection of the agreement quality in mediation, Proc. 3rd International Workshop on Juris -Informatics (Jurisin 2009), pp. 73 - 82 (2009).

[Nitta 09] Nitta, K. et al.: Scenario Extraction Support System Using Word Clustering and Data Crystallization, Proc. 3rd International Workshop on Juris -Informatics (Jurisin 2009), pp. 95 - 106 (2009), (2009).

[Sato 11] Sato, T. et al. : Deliberation Process Support System for Citizen Judge Trial Based on Structure of Factors, Proc. 5th International Workshop on Juris - Informatics (Jurisin 2011), (2011).

[Omoto 11] Omoto, K. et al. : Generation of gesture of embodies conversational agent by using a case base, Proc. Human Agent Interaction Symposium, (2011).